The role of *cis* regulatory domains in behaviour: an evolutionary approach

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by Ursula Paredes-Esquivel

January 2008

Acknowledgements

Many people have helped me in one way or another for conducting my studies. I would like to think first my past and present supervisor, who allowed me to conduct this PhD, which from my distant Peru was almost a dream.

I would like to thank Professor Gabriele Macho, who trusted me and gave me the opportunity to come to the UK, for inspiring in me the interest for the evolution of hominids. I would also like to thank Professor Jim Gallagher who, for allowing me to work and gain invaluable experience (and a living) as a human anatomy demonstrator at the HARC and who whenever I run into trouble always managed to help me.

I would like to thank Professor John Quinn for the infinite amount of patience he has had during these years and for drilling into my hard head the wonders of gene expression. For that, the always-ready moral support and the good humour that kept me on float in my darkest scientific hours, I will forever be in doubt. Special thanks also to Dr. Jill Bubb who, although not on paper, was my "field supervisor" who during all my years as a student always enthusiastically gave me direction and helps with my molecular biology naiveties.

Special thanks to Dr. Sylvia Vasiliou, Dr. Kate Haddley, Mr. Fahad Ali , Dr. Greg Jacobson and Ms Laura Fitton, dearest friends which always kept me smiling and shared hours of fun and hard work that make life at the university so enjoyable. Guys you rock!

To all the past and new members of white block Ms Hannah Stevens, Mr. Stuart Gillies, Dr. Moutasem Abonq, Mrs Lynn Mac Laughlin Dr. Elly Spencer and Mr Iain Kirk thanks for the support, the good humour and willingness to help.

Finally, I want to thank all my loving family in Peru and here in the UK, to those present and those gone, especially to my mother Mrs Carmen Esquivel, my sister Claudia and to Richard, who with endless love and patience kept me sane and always understood me whenever the work got the best of me, for that muchisimas gracias.

Abstract

The molecular processes involved in the evolution of human cognition have been studied thoroughly, revealing little evidence of molecular changes between the brain function of *H. sapiens* other great apes. This suggested that different mechanisms other than in protein or anatomical structure, such as changes in the gene expression of the brain could have mediated the evolution of the human brain. As changes of *cis* acting regulators of gene expression have been documented to mediate many animal phenotypical and behavioural changes I investigated the evolution of the sequences of *cis* acting regulators of the expression of two genes (DRD4 and SLC6A4) that regulate neurotransmission of cognitive and behavioural responses in primates.

The analysis of the evolution of the sequences of the three VNTRs analysed in this study (STin2 and promoter VNTR of the SLC6A4 and the D4ex3 VNTR of the DRD4 gene) demonstrated that the sequences of the *H. sapiens* are markedly different from the sequences of its closest living relative the *P. troglodytes* (common chimpanzee). The differences in the *H. sapiens* and *P. troglodytes* VNTR sequences were tested *in vitro*, by transfecting common VNTR variants into primary cultures enriched for neurons. In these experiments, it was demonstrated that all primate VNTRs supported reporter gene expression, and co-transfection experiments demonstrated that the transcription factors proven to regulate the activities of the *H. sapiens* STin2 and promoter VNTRs also regulated the activity of the non-human primate VNTRs. Most importantly, these experiments demonstrated that the abilities of the *H. sapiens* and *P. troglodytes* VNTRs variants tested to affect reporter gene expression was different. Finally, I investigated the presence and changes in evolutionary conserved domains, proposed to act as important *cis* developmental regulatory domains of the brain. The analysis showed that the ECR1 in the DRD4 gene can act as a *cis* acting regulator of gene expression in cultures derived from early postnatal rat brain; the analysis of the evolution of its sequence shows that the D4ECR1 of *H. sapiens* has accumulated changes after separating from the ancestor hominid as observed for in the VNTRs studies. These observed differences suggest that sometime after *H. sapiens* last shared a common ancestor with other hominids (5 mya), the *cis* regulation of the expression of its SLC6A4 and DRD4 genes diverged, and these changes together with other occurred in different neuronally expressed genes may have contributed to the evolution of the cognitive abilities of *H. sapiens*.

List of Figures

List of Tables

List of Abbreviations

ACC:	Anterior cingulate cortex
ADHD:	Attention deficit hyperactivity disorder
ASPM:	Autosomal recessive primary microcephaly
BL:	Branch length
BLAT:	BLAST-like alignment tool
bp:	Base pairs
cAMP:	Cyclic AMP
cc:	Cubic centimetres
CCR5:	Chemokine receptor 5
ChIP:	Chromatin immunoprecipitation
CMAH:	CMP-sialic acid hydroxylase
CNS:	Central nervous system
CSD:	Cold shock-domain
CpG:	CG dinucleotides
CYP7A1:	Family 7, subfamily A, polypeptide 1
DA:	Dopamine
DAT:	Dopamine transporter protein
DMEM:	Dulbecco's modified eagle's medium
DMSO:	Dimethylsulphoxide
DNA:	Desoxi ribonucleic acid
dNTP:	Deoxynucleotide triphosphate
DPE:	Downstream promoter element
DRD2:	Dopamine receptor D2
DRD4:	Dopamine receptor D4
EBI:	European bioinformatics institute
ECR:	Evolutionary conserved region
EDTA:	Ethylenediamine tetraacetic acid
EMSA:	Electrophoretic mobility shift assay
FCS:	Fetal calf serum
FOXP2:	Forkhead box P2
GTFs:	General transcription factors
HAT:	Histone-acetyltransferases
HDACs:	Histone deacetylases
HIV:	Human inmunodeficiency virus
HOM-C:	Homeotic C
Hox:	Homeobox
Inr:	Initiator
INS:	Insulin
IPC:	Information processing capacity
IPTG:	Isopropyl β-D-1-thiogalactopyranoside
IRBP :	Interstitial retinoid binding protein
kb:	Kilobases
LiCl:	Lithium chloride
LDLR:	Low density lipoprotein receptor
l/s:	Long and short variants of the 5' promoter VNTR of the SLC6A4 gene
LB:	Luria-bertani broth
MAOA:	Monoamino oxidase A

MeCPs:	Methyl-CpG-binding proteins
MDR1:	Multidrug resistance 1
mRNA:	Messenger RNA
MW:	Molecular weight
Mya:	Million years ago
nAChR β4:	Neuronal nicotinic receptor subunit β4
NCBI:	National center for biotechnology information
NOS-I:	Nitric oxide I
OD:	Optical density
Otx2:	Orthodenticle homeobox
PATPAL:	Patronato del parque de las leyendas
PCR:	Polymerase chain reaction
PDYN:	Precursos of dynorphin
PPTA:	Preprotachykinin-A
PATPAL:	Patronato del Parque de las Leyendas
REST:	Repressor element 1-silencing
RNA:	Ribonucleic acid
RPMI:	Roswell Park Memorial Institute
RT-PCR:	Reverse transcriptase PCR
SLC6A3:	Solute carrier family 6 (neurotransmitter transporter, dopamine), member 3
SLC6A4:	Solute carrier family 6 (neurotransmitter transporter, dopamine), member 4
SN:	Substantia nigra
SNP:	Single nucleotide polymorphisms
SOC:	Super Optimal Broth for catobolite repression
Sp1:	Specificity factor 1
SREBF1:	Sterol regulatory element-binding protein
SV40:	Simian virus 40
TAFs:	TBP associated factors
TBE:	Tris Borate EDTA.
TBP:	TATA binding protein
TF:	Transcription factor
TFBS:	Transcription factor binding sites
TK:	Thymidine kynase
TR:	Tandem repeat
tRNAs:	Transporter RNA
UCSC:	University of California Santa Cruz
UTRs:	Untranslated regions
UV:	Ultraviolet irradiation
VNTR:	Variable number of tandem repeat
VTA:	Ventral tegmental area
WL:	Wave length
X-gal:	5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside
YB-1:	Y-box biding protein
5HT:	Serotonin
5HT1A:	Serotonin 1A receptor
5HT2A:	Serotonin 2A receptor
5HTT:	Serotonin transporter protein

Contents

page number(s)

Chapter 1 Introduction

1.1 Evolution of the hominid brain

Two hallmarks have marked the evolution of genus *Homo*, the first is the evolution of bipedalism, considered by many authors as an adaptation to new habitats and to different food resources (Foley and Lee, 1991; Joffe and Dunbar, 1997; Vaughan, 2003) and the second is the remarkable enlargement and increased complexity of the brain (Rightmire, 2004). This second hallmark is considered to have defined modern humans and be responsible for the cognitive prowess of the modern human mind.

The most dramatic changes in the brain of lineages leading to modern *Homo sapiens* started approximately 2.2 million years ago (mya) by the end of the Pliocene and beginning of the Pleistocene (Gervais and Wilson, 2005). The brain enlargement is apparent in the remains of fossil skulls from *Autralopithecus africanus*, *A. robustus*, *Homo erectus*, *H. habilis* and *H. sapiens* (Figure 1.1). During this period, the cranial capacity of our ancestors expanded from 500 cc as seen in *A. africanus* (2.2 mya), a size comparable to the brain of modern apes, to 700 cc as seen in *H. habilis* (2- 1.6 mya) and later to 1060 cc as seen in *H. erectus* (1-0.7 mya). The second greatest enlargement of the brain occurred with the first appearance of *H. sapiens* (130 000 years ago) increasing from 1040 to1595 cc. At this time, there was also a reduction of body mass, which caused a relative increase of brain size (Kappelman, 1996). This later change in brain/body size proportion coincided with the greatest expansion of the population of *H. sapiens* around the globe (Kappelman, 1996).

The increase in brain size and the morphological changes accompanying the evolution of bipedalism appeared on the fossil record of *Homo* simultaneously; therefore they have been proposed to be entwined (Rosenberg and Trevathan, 2002).

The evolution of bipedalism transformed the primitive human pelvis, forcing the newborn to make his way not downward but forward, through a narrow and inextensible pelvic outlet (Houdart, 2005). Although it is not known whether one evolutionary event drove the other, the anatomical change of the pelvis is a constraint for the maximum size of the newborn head at birth, and it is believed that this pelvic transformation is linked to the secondarily altricial characteristic of the human brain (Rosenberg and Trevathan, 2002; Shanley and Kirkwood, 2001). This means that in modern humans, the development of the brain does not cease shortly after birth (as seen in monkeys and apes), but undergoes dramatic changes in response to hormonal, physiological and environmental cues until the onset on adulthood. The changes are more pronounced during adolescence, a period that is considerably longer in humans than in the great apes (Varki and Altheide, 2005) and that is characterised for extensive remodelling of the brain (Sisk and Foster, 2004). This process includes increased myelination, decreased grey matter volume in cortical areas, changes in connectivity in the amygdala and frontal cortex.

The remarkable transformation undergone by the human brain in the last 2 million years and its responsiveness to external and internal stimuli demonstrates its intrinsic plasticity, which is the key mechanism for adaptation, development and learning, as much as a cause of pathology. This plasticity has produced heritable changes, which constitute the basis for human brain evolution. Where in the brain these changes have occurred, why they have occurred and most importantly, what are the mechanisms for this heritable plasticity are questions for constant debate. In the following sections, I present a summary of the comparative anatomical and genetic studies and their most relevant findings in the field of human cognitive evolution, which inspired my thesis.

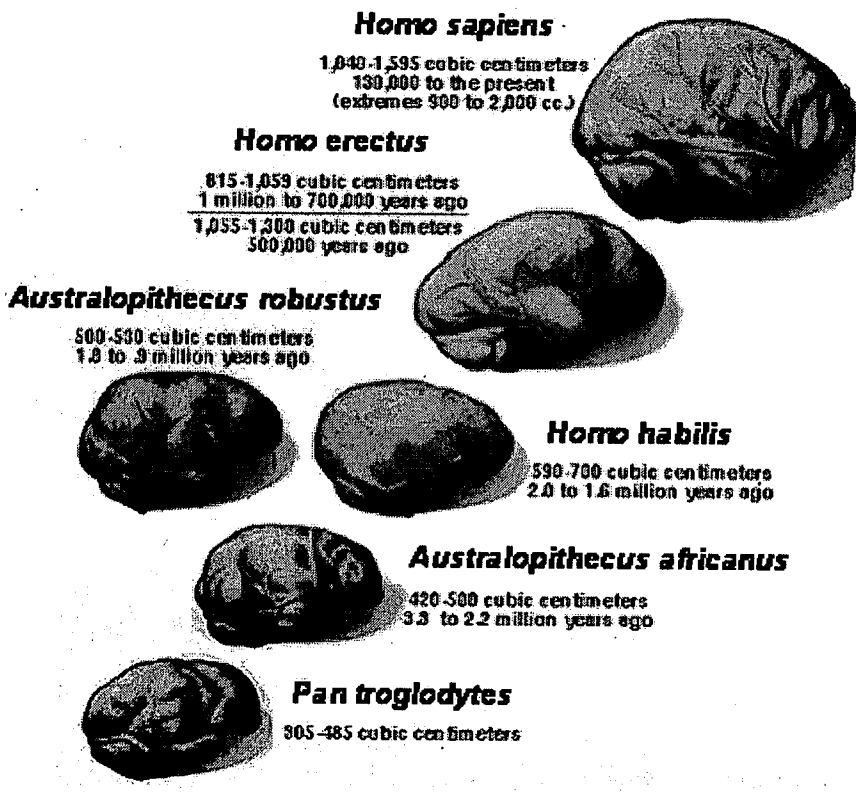


Figure 1.1 Enlargement of brain during the evolution of the Hominin (which include the genera: *Homo* and *Australopithecus*) lineage (Extracted from: www.evolution.massey.ac.nz). A diagram of the *P. troglodytes* brain is included for comparative reasons.

1.2 Anatomical and histological differences between modern human and non-human primate brains

At the start of the 20th century, Brodmann (1909) compared human and non-human primate brain anatomy and histology. His studies suggested that in comparison to old world monkeys, modern human brains presented an overdeveloped frontal cortex (reviewed by Allman 2002). Given the importance of the frontal cortex in high executive function (e.g. decision making, memory, social behaviour), it was believed that the cognitive differences between modern humans and great apes originated from this enlargement (Semendeferi et al., 1997). For many years Brodmann's findings remained unquestioned; however, recent allometric studies using more appropriate comparison groups (great apes instead of old world monkeys) and increase in sample number have demonstrated that the size of the modern human frontal cortex is not bigger than that expected for a hominid brain of its size (Semendeferi et al., 1997). However, these studies did confirm that the frontal cortex is larger in hominids (modern humans and great apes) than in other primate and non-primate species. Nevertheless, the development of the hominid brain has been encompassed by the evolution of a few specializations in the cerebral cortex. Indeed, the anterior cingulate cortex (ACC) is involved in the regulation of autonomic responses and diverse cognitive functions such as self-awareness; the discrimination of information from conflicting cues, focused problem solving and error recognition (Allman et al., 2001; Posner and Rothbart, 1998; Wenderoth et al., 2005). This region presents 2 specialised cellular types unique to hominids located in layer V (Figure 1.2), a layer which typically relays the output of cortical processing to other cortical areas and subcortical structures. One of these is a group of large bipolar spindle neurons whose axons are known to project into the underlying white matter (Nimchinsky et al.,

1999). In hominids, these neurons are significantly different from others in layer of V, because their volume is four times greater than the average layer V pyramidal neuron. Although a given function has not been attributed to these cells, it is known that they are severely affected by degeneration in Alzheimer's disease (Nimchinsky et al., 1999). The second specialised cellular type is a neuronal population of pyramidal neurons that contain the calcium-binding protein calretinin. This protein may act as an ion calcium buffer in neuronal populations and are often colocalized with GABA in non-pyramidal neurons, however its role in this pyramidal neurons is poorly understood. In non-hominids this protein is present in bipolar neurons in superficial layers of the ACC whereas in hominids this protein is found in pyramidal neurons of layer V (Hof et al., 2001). Both cellular types are more densely distributed in the ACC of modern humans than in that of great apes (Allman et al, 2002). Another specialization of the hominid brain is the volume increase of Brodmann's area 10, located in the frontal pole (Figure 1.2). In comparison to great apes, this area is larger, both absolutely and relatively in modern humans. However, unlike the spindle and pyramidal cells, area 10 is not unique to hominids but is present as a much smaller region in lesser apes and monkeys (Semendeferi et al., 2002). It has been proposed that as indicated by the morphology and extension of their axons, these specialised neurons have evolved to convey information originating in ACC to other parts of the brain, especially with the co-evolving Brodmann's area 10 (Allman et al., 2001). Based on functional MRI studies that suggest a the role of the ACC in decision-making, communication and cognition, Bush (2002) proposed that the described ACC specializations may have evolved to process emotional and cognitive behaviour which originate and required in a social environment.

Another hypothesis to explain the cognitive evolution of hominids proposes that the exceptional mental abilities of modern humans may be a result of functional rather than gross anatomical evolution (Semendeferi et al., 1997). For example it has been observed that the connectivity of the neurons could increase without the necessary increase in volume of a given brain area. Indeed, neuronal connectivity has been directly related to cognitive performance in aging brain and neurological disorders in humans and animal models (Bussière and Hof, 2004; Mahncke et al., 2006; Morrison and Hof, 1997). This functional evolution hypothesis would provide a better explanation to the markedly different behavioural outputs in some brain regions (e.g. Broca's area of language) which do not exhibit organizational or morphological differences across hominids (Sherwood et al., 2003). Indeed, human cognitive abilities are thought to be a product of a combination of the largest number of cortical neurons and greatest information processing capacity (IPC), determined by diameter of myelinated fibres (Roth and Dicke, 2005), and not simply by the total or relative volume of the brain.

In summary, all hominid brains posses the same specializations but these are more developed in modern human than in great apes. However, the reason why modern humans experienced greater development of these specializations and cognitive functions than the great apes remains unanswered. This motivated the formulation of two hypotheses: the social cognition and the cognitive ecology hypotheses that aim to explain the environmental conditions that could have contributed to the cognitive evolution of modern humans. These hypotheses are summarised in the following section.

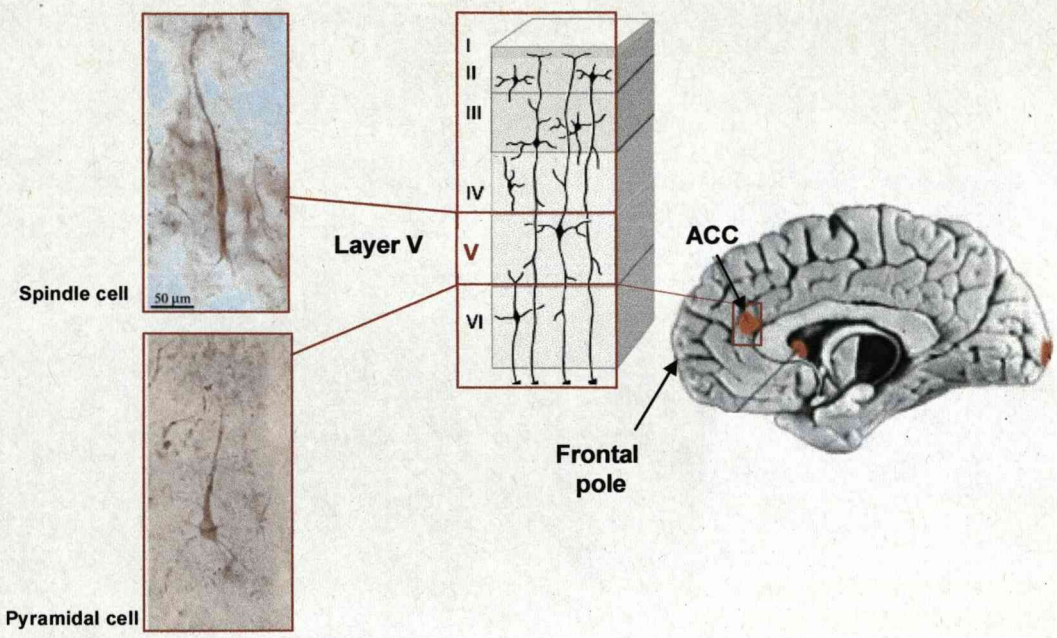


Figure 1.2 Layer V of the anterior cingulate cortex (ACC) contains two specializations of the hominid brain. In the brain diagram, the ACC is marked by a red rectangle. In the hominids, this area presents spindle neurons and calretinin positive pyramidal neurons. Another specialization, Brodmann's area 10, has been located in frontal pole. Modified from Khaitovich (2004b) and Allman (2002). Medial views of the *H. sapiens* brain.

1.3 Evolution of cognitive abilities of the hominid brain

In 1871 Charles Darwin classified modern humans as members of the primate order and proposed the "descent of humans from apes", receiving heavy criticism by the scientists of the time. After the theory of human evolution was widely accepted, the scientific community concentrated in identifying the traits that define human nature and differentiated modern humans from other primates. Apart from the obvious anatomical traits which differentiate modern humans from great apes e.g. perlvic morphology, body hair, nostrils orientation; there are also several cognitive traits that for many years were acknowledged to be exclusive to modern humans (e.g. enhanced learning abilities, speed of information processing, syntactical and grammatical language, numeracy, artistic expression (Varki and Altheide, 2005). More recently, it

has become clear that although less developed, some of the mentioned cognitive capacities are also present in the great apes (Collier-Baker et al., 2006; Matsuzawa, 2007).

Since the 1990's two schools of thought emerged to explain the evolution of cognitive abilities of modern humans. The first (the social brain hypothesis) proposed that in a complex social environment, more processing speed would be required to maintain social status and relationships (Dunbar, 2003; Dunbar, 1992). In turn, this would drive the greater development of brain function thus producing a correlation between the size of the "neocortex" (outer cortical layer in the cerebral hemispheres of the mammalian brains, Figure 1.3) and the size of social groups formed by primates (Figure 1.4). Such correlation is greater when group size is compared to orbitofrontal cortex, an area which processes visual information and plays a key role in the interpretation of information from primate groups (Barton, 1996; Joffe and Dunbar, 1997).

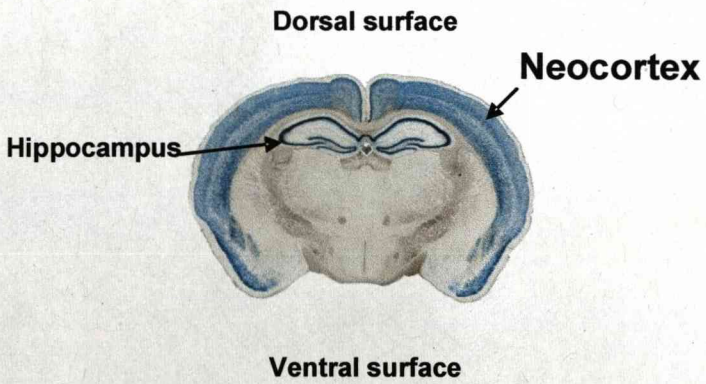


Figure 1.3 Mamamlian neocortex. LacZ-staining on a frontal brain section of a NEX-CRE LoxP-LacZ-indicator mouse showing staning of the hippocampus and Neocortex (extracted from the Department of Neurogenetics of the Max Planck Institute of Experimental Medicine webpage: http://nave.em.mpg.de/nex_cre/)

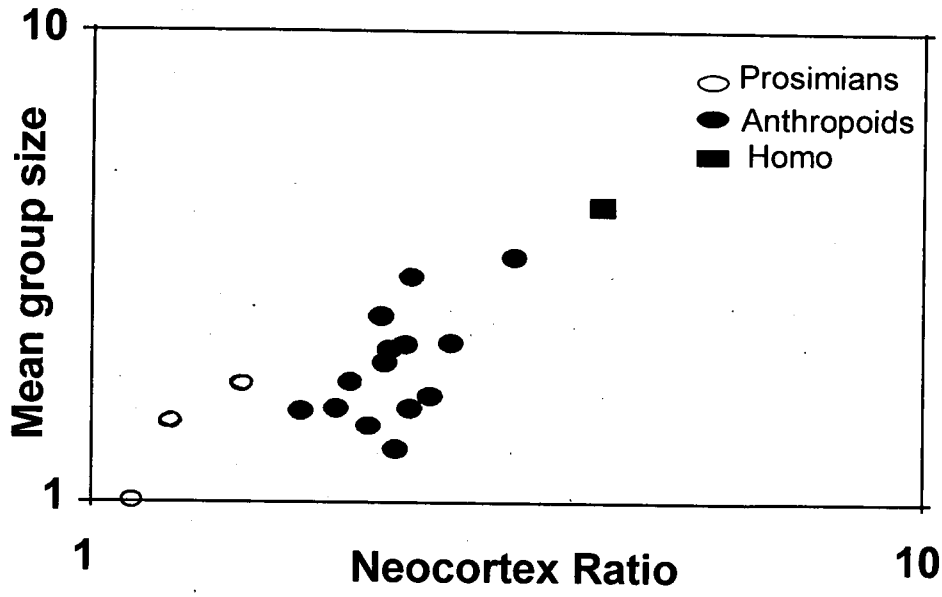


Figure 1.4 Neocortex area increases with the size of the social group. The volume of neocortex is greater in anthropoids than in prosimians, and is greatest in modern humans (modified from Barton, 1996).

A second hypothesis (cognitive ecology) proposes that adaptation to new ecological conditions played the most important role in the evolution of hominid brain (Foley and Lee, 1991). Amongst the many ecological factors postulated to have contributed towards the encephalization of *H. sapiens* is the adaptation of foraging behaviour (Barton, 1998). These authors suggested that hominids changed their diets to one more rich in fruits, animal protein and fat and that the extraction, capture and mapping of resources could have contributed to the enlargement of the modern human brain. This hypothesis was supported by studies on the degree of "frugivorism" presented by primates and bats, observed to correlate with volume of their neocortices (Jones and MacLarnon, 2004). This has been further corroborated by the correlation between complex distribution of food resources, greater memory and enlargement of the cerebellum found in mammals and other vertebrates (Healy and Hurly, 2004). The cognitive ecology hypothesis also suggest that "group foraging" may have contributed

to the reduction of energy invested in capture of prey, favoured the formation of social groups and the formation of modern human societies (Foley and Lee, 1991). In recent years it has become more evident that ecological and social factors may have contributed equally to the evolution of human cognition; therefore, the two cognitive hypotheses described above should be regarded as complementary and not mutually exclusive explanations for the evolution of the modern human brain (Reader and Laland, 2002).

Evolutionary biologists studying human cognitive evolution explored different avenues, which include the genetic elements that may have evolved over the course of human evolution, perhaps contributing towards the development of the function of the human brain. In the following section, a synthesis of molecular, biochemical and cellular studies of human cognitive evolution is summarised.

1.4 Biochemical, cellular and molecular studies of human evolution

Since the turn of the 20th century, scientists have tried to address the molecular and genetic basis of human nature. These methods include the comparison of proteins and DNA sequences by electrophoretic, immunological, and sequencing techniques. One of the first cellular studies of the evolution of hominids was conducted in 1904 by Nuttall and colleagues, who used serological cross-reactions to study the phylogenetic relationships amongst primates (reviewed by Goodman, 1967). Their study confirmed morphology-based studies, which suggested that modern humans were more closely related to African than to Asian great apes. In the second half of the 20th century, other cellular tools such as chromosome banding were incorporated to the study of human origins. Dutrillaux (1980), Turleau and de Grouchy (1972), and others conducted the first comparative analyses of hominid karyotypes. They found that rearrangements such as pericentric inversions and chromosome fissions had occur

during evolution of hominids genomes. However, they also demonstrated that the basic organization of primate chromosomes was highly conserved across primates. Although differences in cytogenetic organization between humans and apes were observed, these studies concluded that the degree of similitude amongst hominid chromosomes was not sufficient to explain differences amongst hominid species (Turleau and de Grouchy, 1972).

The arrival of protein electrophoresis allowed the first comparisons of protein sequences. These studies demonstrated that all structural proteins studied e.g. alpha, beta and gamma globins, fibrinopeptides and cytochromes (Boyer et al., 1971; Palmour et al., 1980; Sullivan, 1971) of hominids shared 99-100% similarity in peptide sequences. When DNA sequencing became available, scientists were able to compare the DNA coding sequences of so-called "house keeping genes" of hominids (Garver and Talmage, 1975; Goodman et al., 1971; Sarich and Wilson, 1967). These were first isolated because they have a key role in regulating basic cell functions, thus are ubiquitously expressed and highly conserved across species. In these initial genome comparisons, the exonic sequences were predominant as the mRNA was easier to prepare and analyse. The comparisons demonstrated that although there was greater differentiation between modern humans and great ape DNA sequences than the observed in peptide sequences, these nucleotide changes were mainly positioned in synonymous locations. This further corroborated that little change had occurred in proteins and their coding DNA sequences for 12-21 million years, since hominids last shared a common ancestor (Salem et al., 2003). The degree of similarity found amongst hominids protein sequences reflects the slow rate of exonic evolution and failed to explain the majority of phenotypic distinction that exist amongst modern hominids.

In spite of this little variation found in the protein sequence comparison, there are a few examples of how single mutations within proteins (caused by a non-synonymous nucleotide substitution) could have directly contributed to the evolution of the modern human brain and its cognitive capacities. For example these include mutations within the monoamino oxidase A (MAOA), autosomal recessive primary microcephaly (ASPM) and forkhead box P2 (FOXP2) genes, involved in neuronal growth, behaviour, language and brain size (Andres et al., 2004; Kouprina et al., 2004; Zhang et al., 2002). Furthermore, it has been suggested that these and other brain-expressed genes have undergone a faster rate of evolution in modern humans than in great apes. This faster rate of evolution has been interpreted as a sign of strong selective pressure in some neuronal genes on the lineages leading to modern humans (Dorus et al., 2004). However, recent analyses using more appropriate out-groups (rhesus macaques instead of rodent sequences) dispute this hypothesis, suggesting that no greater speed of the evolution of genes expressed in the modern human brain exist when compared to genes expressed in the brain of *P. troglodytes* (Bakewell et al., 2007). It has been suggested that these discrepancies arise because only recently scientists have considered that the differences found between modern humans and great apes genomes could merely reflect silent or adaptive changes in the ape genomes, with no implications on the evolution of modern humans' nature (Bakewell et al., 2007).

The use of intronic and intergenic DNA sequences in the study of the evolution of hominids only began in the 1980's (Scott et al., 1984; Slightom et al., 1985). This was because incorrectly, these regions were thought to be devoid of "functionality" (Ohno, 1972). It was not until the mid 1980's and 1990's that scientists began to understand the complexity of eukaryote genomes. They realized that apart from the proximal

promoter (which can be located hundreds of bases upstream of the transcriptional start site); many regulatory elements of the gene expression machinery (enhancers, repressors) were located in intronic and intergenic regions (Battersby et al., 1996; Conrad and Botchan, 1982; Gillies et al., 1983). The completion of the *H. sapiens*, *P. troglodytes* and *Mus musculus* (laboratory mouse) genome projects at the end of the 20th century has permitted genome wide comparisons. This demonstrated the presence of sequences equally (or more) conserved than exons and proximal promoters, which are located in intergenic or intronic regions. This conservation suggests an important regulatory role for these elements (Nie et al., 1996; Nishizaki et al., 2001; Davidson et al., 2006, Prabhakar et al., 2006b). Furthermore, the inclusion of non-coding DNA sequences in the study of human evolution has broadened the genetic distance that exist between modern humans and the great apes (Chen et al., 2001). Therefore, it is possible that the regulatory elements that might be responsible in part for the evolution of the functional phenotype of the modern human brain exist in intronic and intergenic regions.

The lack of evidence supporting a solely protein-based evolution of the modern human brain and the discovery of regulatory domains in non-coding regions motivated the formulation of two hypotheses, which aim to explain the evolution of the modern human phenotype. These are the "changes in gene expression" (King and Wilson, 1975) and the "loss of gene function" (Olson et al., 1999) theories, which are discussed in section 1.4.1.

1.4.1. Theories of genetic evolution of the modern human phenotype

King and Wilson (1975) proposed that changes in regulation of gene expression are the key factor in creating differences between modern human and great ape phenotypes (physiological, behavioural and anatomical). This phenotypic

variation is caused by differences in levels of expression at which protein products of the same genes are produced in specific tissues. Such variation in levels of gene expression would be mediated by the presence of polymorphisms in the promoter or control regions of genes or by chromosomal rearrangements. This hypothesis was tested almost 3 decades later (Cáceres et al., 2003; Enard et al., 2002a). These authors independently compared the transcriptomes of *H. sapiens* and *P. troglodytes* brains and other organs (Figure 1.5). These two groups showed that in comparison to the expression profile of genes in other organs (e.g. liver or heart), the genes expressed in the brain were most different between of these two species (Figure 1.5). Later, Khaitovich (2004b) demonstrated that there was no great intraspecific variability in the transcriptomes of the cerebral cortex of *H. sapiens* and *P. troglodytes* (n=5 for both species). However, they also found that, the expression patterns of different regions such as the cerebral cortex, the caudate nucleus and the cerebellum differ greatly within each individual. The authors also analysed samples from the ACC from *H. sapiens* and *P. troglodytes*, shown to present cellular differentiation between the two species. However, they were unable to identify gene expression changes in the ACC. Finally, they discovered that at least 10% of genes expressed in a given region of the brain are differentially expressed in *H. sapiens* and *P. troglodytes*. Heissig (2005) further investigated the *in vitro* function of proximal promoters (1kb upstream and 0.5 kb downstream) of genes that were found diferentially expressed in the brain cortex of *H. sapiens* and *P. troglodytes* using microarray technology. This *in vitro* analysis showed that 7 of 12 proximal promoters analysed supported differential reporter gene expression *in vitro* in a human neuroblastoma (SHEP 27) and in a cervix carcinoma (C33A) cell line.

Given that the above described findings are based on a small sample size, and as the neuronal patterns of gene expression in one individual, population and species is highly variable during different developmental stages, seasons and even during circadian cycles (Barrett et al., 2006; Sisk and Foster, 2004 ; Watanabe et al., 2007), these results should be interpreted with caution. However, in spite of these caveats, these findings suggest that differences in the gene expression between the modern human and great ape brains do exist, and this might involve differential regulation of the transcriptional machinery of neuronally expressed genes.

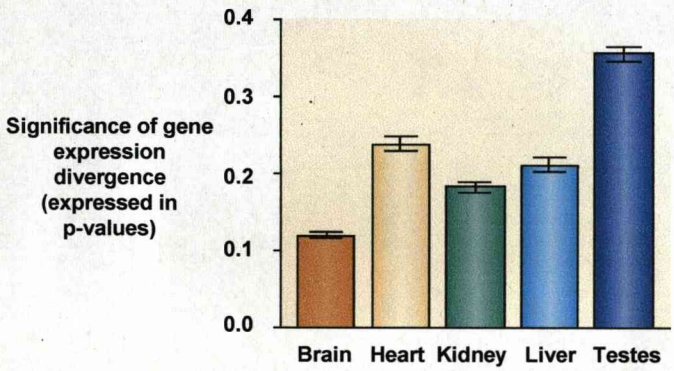


Figure 1.5 Significant differences between the levels of gene expression found in different organs of *H. sapiens* and *P. troglodytes*. Expression levels of different organs were compared. The divergence (expressed in significance or *p*-values, where error bars show 95% confidence intervals calculated by 1000 bootstraps over all genes expressed in a tissue) of levels of gene expression between the two species is greater in the brain (in orange) than that found between kidneys, heart, liver and testes (modified from Khaitovich et al., 2006)

A second hypothesis to explain the evolution of human phenotype was formulated by Olson and colleagues (1999). The "less is more" hypothesis proposes that during speciation processes such as that undergone by the human lineage, individuals are exposed to different selective pressure imposed by the new nvironment. Thus, loss of pre-existing traits and consequently the genes that produce such traits can occur. This loss of gene function would not create deleterious effects

on the fitness of an organism on the new environment; then, these genes would not be lost and remain in the genome to be eventually used if a new shift in selective pressures would favour it. Examples of this model of phenotypic change are the acquired resistance to *Plasmidium falciparum* (parasite which causes malaria) which has correlated with the loss of function of the Duffy gene (that encodes for the erythrocyte chemokine receptor) in humans descendent from West African populations (Escalante et al., 1995). Other notorious examples are the loss of function of the chemokine receptor 5 (CCR5) gene, which produces resistance to infection by the HIV retrovirus (Galvani and Novembre, 2005); the inability of humans to digest Sialic acid, caused by the inactivation of the CMP-sialic acid hydroxylase (CMAH) gene and the loss of function of genes involved in olfaction (Varki and Altheide, 2005). The validity of this hypothesis as the main mechanism for phenotypic change in humans is debatable, since studies in support of the "less is more hypothesis" are scarce. Although the loss of gene function is thought to occur more regularly than divergence of gene expression profiles and proteins function (Olson and Varki, 2004), this mechanism has been shown to occur in genes that form part of large families, where it is possible that other closely related genes could overlap in function (Olson and Varki, 2003). Moreover, it dismisses the importance of duplications of individual genes, chromosomal segments, or entire genomes as important mechanisms for species diversification (Lynch and Conery, 2000). Due to these deficiencies, many scientists have embraced the gene expression hypothesis of King and Wilson to explain the evolution of human cognition and phenotype. The major mechanisms of regulation of the expression of a gene, from DNA to mature protein in the cell, are briefly discussed in the following sections.

1.5 Mechanisms of modulation of gene expression

The fate, physiology and function of a cell are specified by the genes expressed in it. The regulation of gene expression can be affected by metabolism, the presence of stimuli and developmental stage of the organism (Gerrard et al., 2005; Sabatini et al., 2007; Zvonic et al., 2007). The levels of gene expression of a protein-coding gene can be modulated during all the stages that exist between the DNA and the production of protein (Jaenisch and Bird, 2003; Latchman, 1992). These mechanisms include chemical and structural modification of chromatin (acetylation, deacetylation) and DNA (methylation); regulation of efficiency of transcription of the DNA into the RNA transcript, modification of the RNA transcript by post-translational processes (e.g. alternative splicing, capping, polyadenylation, RNA degradation). In recent years other mechanisms of gene expression regulation which involved the activity of two clases of RNAs, micro RNA (miRNA) and small interfering RNAs (siRNA) have been identified. Both can inhibit protein translation thus producing repression of gene expression mediated by their ability anneal to mRNA, although they have been suggested to also affect protein translation (Pillai et al., 2007). The initial stage of transcription plays a major role in gene expression because it is then when it is decided which genes will be transcribed into the primary RNA transcript. In the following sections the formation of the basic transcriptional machinery and some common mechanisms for modulation of gene expression in eukaryotes are briefly discussed.

1.5.1 Basal transcription

The process by which a DNA sequence is transcribed into a RNA transcript is known as transcription. In eukaryotes, three different types of RNA polymerases are responsible for transcribing different subsets of genes. The RNA polymerase I

transcribes genes encoding ribosomal RNA; the RNA polymerase II transcribes genes encoding mRNA and certain small nuclear RNAs and the RNA polymerase III transcribes genes encoding tRNAs and other small RNAs (reviewed by Novina and Ananda 1996). Transcription directed by RNA polymerase II is divided in two phases: (1) initiation of transcription which involves the recruiting of RNA polymerase II and associated proteins (around the core promoter), to produce the copy of DNA and (2) the synthesis and processing of RNA.

The RNA polymerase II is not capable of specific transcriptional initiation by itself, but needs the formation of a complex of general transcription factors (GTFs, such as: TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH see Figure 1.6) (Orphanides et al., 1996). Only when these are assembled on the DNA promoter, can transcription start. The minimal promoter for the RNA polymerase II is defined as the set of DNA sequences required for assembly of the pre-initiation complex (formed by GTFs). Transcription initiated by this minimal set of proteins (as seen in Figure 1.6) is termed basal transcription (Latchman, 1993).

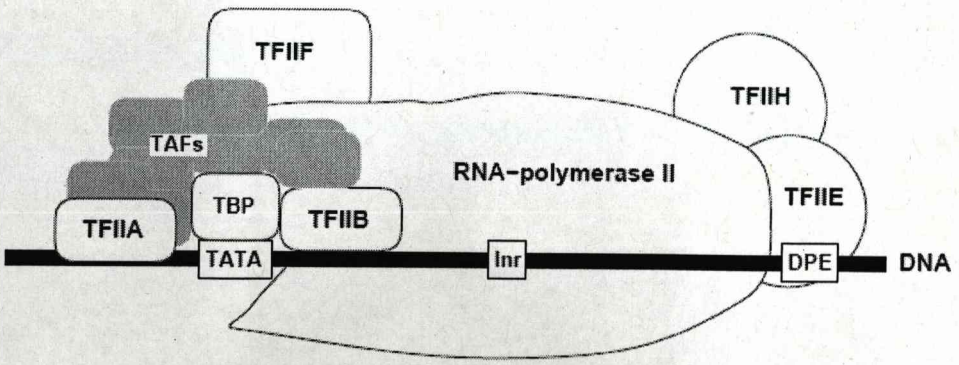


Figure 1.6 Core promoter associated with RNA polymerase II and GTFs. Core promoter elements shown are the TATA-box (TATA, usually around -35-25 bp), the initiator element (Inr, around the start point), and the downstream promoter element (DPE, around +30).

Once assembled, the initiation complex is phosphorylated and the polymerase is able to leave the complex and start synthesizing RNA. The list of the proteins involved in this complex continues to grow, but the major proteins involved are shown in Figure 1.6.

There are different types of minimal or core promoters. One commonly found class consists of a TATA-box (consensus TATAAAA) located around 25-35 bp from the transcription start site, which directs transcriptional initiation (White and Jackson, 1992). The TATA box is bound by the TATA binding protein (TBP), a subunit of the TF TFIID (Parvin and Sharp, 1993). The TFIID also contains a number of TBP associated factors (TAFs) which mediate the attachment of the TBP to the core promoter. The exact position of the transcriptional start point is often determined by another element, the initiator (Inr) (Smale and Kadonaga, 2003). The Inr was defined as a discrete core promoter element that does not have a very strict consensus (YYCAYYY, where Y= pyrimindine rich), it is functionally similar to the TATA box and can function independently of it. Additional to these types of promoters, there are also those that have both TATA and Inr sequences and promoters that do not have either (Pugh and Tjian, 1991; Smale, 1997). Another promoter element is present in some TATA-less, Inr-containing promoters about 30 bp downstream of the transcriptional start point (Burke et al., 1998). This element, which is known as the downstream promoter element (DPE, Figure 1.6), appears to be a downstream analogue of the TATA box in that it assists the Inr in controlling precise transcriptional initiation.

The tissue and cell specificity of gene expression is driven by the recognition of a TF to the binding sites present in the promoter region or those found in enhancer or repressor domains. These domains can be located at varying distances from the

transcriptional start site, some can lie adjacent to the core promoter, while others can be located several tens or hundreds of kilobases upstream or downstream of the promoter. These sequences exhibit binding sites for TFs that increase or repress the level of transcription from core promoters (Wasylyk, 1988; Zawel and Reinberg, 1993). The arrangement of different binding sites in the regulatory domains and the availability of specific TFs, which can recognise binding sites generates functional specificity (reviewed in He and Rosenfeld 1991; Quinn, 1996). As mentioned previously, TFs often require the binding of other TFs (either ubiquitously present or stimuli inducible) to the regulatory domain or the presence of TF with which they form hetero or homo duplexes or protein complexes which further increase the specificity of recognition of the binding sites for the regulation of gene expression. Regulatory domains can also act as developmental regulators (Geyer and Corces, 1987), which mediate gene expression only active during specific developmental stages.

Transcriptional regulation is regulated by the structural arrangement of TFs binding sites within DNA and the arrangement of associated histone proteins (nucleosome complex). Moreover, the interaction of TFs bound elsewhere in the DNA sequence with those in the initiator complex most likely mediate their affects via combinatorial actions on nucleosomes and direct interactions with the basal transcriptional machinery. The organisation of DNA in chromatin, nucleosome formation, nucleosome modifications and their effects on transcription are discussed in the following sections.

1.6 Chromatin modification and gene expression

Eukaryotic DNA is packaged into chromatin. It is becoming increasingly clear the transcription of genes is regulated, at least in part, by chromatin-remodeling

events, which modulate the accessibility of TFs to the DNA tightly packaged into chromatin (Hull, 1982). The basic structural organization of DNA in chromatin of the eukaryotic cell is the nucleosome core plus linker DNA (Turner, 1991). The nucleosome core consists of an octamer formed by two copies each of histones: H2A, H2B, H3 and H4 plus 146 base pairs of DNA organized in 1.75 turns around the histone octamer (Cress and Seto, 2000). The bulk of nucleosome core particles are associated with histone H1 (linker histone). The nucleosome core plus histone HI and the entire linker DNA is called the nucleosome (Klug et al., 1980; Kornberg and Klug, 1981). Many studies have revealed that modifications of the amino terminus of the histones in nucleosomes play important roles in the regulation of gene expression. Histones control gene expression by modulating the structure of chromatin and therefore, accessibility of regulatory DNA sequences to TFs (Travers 1987; Verdin et al., 2003). Mechanisms of modification of nucleosomes include phosphorylation, methylation, ubiquitination and acetylation, all which have been correlated with gene activation or repression (Lusser, 2002; Shilatifard, 2006). Acetylation and deacetylation of histones has been thoroughly investigated and will be briefly discussed in section 1.6.1.

1.6.1 Acetylation and deacetylation of DNA

Histone acetylation is produced by the attachment of an acetyl group to the lysine amino acids of the N terminus of each histone by histone-acetyltransferases (HATs). Acetylation occurs at specific lysine residues, all of which occur in the amino-terminal domain of the core histones. Histone acetylation is potentially a major influence on events such as transcription, replication, DNA packaging and DNA repair. In transcription, this chemical modification of histones allows the access of TFs (Ng and Bird, 1999). Each acetyl group added to a histone reduces its net positive

charge by 1, neutralizing its charge. This reduces the interaction of the N terminus of histones with the phosphate groups of DNA. As a result, the condensed chromatin is transformed into a transiently relaxed structure, which allows genes to be transcribed (Eberharter and Becker, 2002). The transformation of histones (and chromatin structure) via acetylation could occur in several ways. For example, acetylation can intervene in the processing and deposition of displaced histones during the formation of nucleosomes with newly synthesized histones, it can also affects the reformation of nucleosomes core particles from displaced histones and the reassembly of the higher order structure of chromatin (Csordas, 1990).

A mechanism that reverts the gene activation caused by histone acetylases is histone deacetylation, conducted by histone deacetylases (HDACs) (Wang et al., 2004a). HDACs are chromatin-remodelling factors that deacetylate histones by catalyzing the removal of acetyl groups on the amino terminal lysine residues of the core histones. Recruitment of HDACs to a promoter can result in localized histone deacetylation and leads to transcriptional silencing or repression. In modern humans, at least sixteen different HDACs have been reported, being classified based on sequence identity and domain organization. HDACs and are sub-divided into class I (HDAC 1, 2, 3, 8), class II (HDAC 4, 5, 6, 7, 9 and 10), class III (homologous to sirtuins proteins found in the yeast SIRT 1 to 7) and class IV (HDAC11). There is evidence that these different HDACs target different patterns of acetylation and regulate different genes (Verdin et al., 2003). An example of such a protein that regulates neuronal gene expression is the repressor element 1-silencing TF or REST, that mediates repression through histone-deacetylase, histone demethylase and histone-methylase activities (Ooi and Wood, 2007).

1.7 DNA chemical modification

In eukaryotic genes cytosine base can be changed into 5-methylcytosine by the addition of methyl groups by enzymes known as DNA methyltransferases e.g. Dnmt1, Dnmt3a and Dnmt3b (Turek-Plewa and Jagodzinski, 2005). This methylation pattern is restricted to the cytosines in CG dinucleotides (CpG), which results in gene silencing or repression (Bock et al., 2006). CG dinucleotides are often enriched in islands (compared with average of the genome). These islands are often but not solely found in the 5' region of a gene are often associated with gene regulation properties. For example, constitutively expressed "housekeeping genes" have unmethylated CpG islands, and tissue specific genes are found to remain unmethylated preferentially in tissues where they are expressed (Bird, 1986). Methylation is believed to influence gene expression via interactions with Methyl-CpG-binding proteins (MeCPs). The MeCPs protein form part of histone deactylase (HDAC) complexes that attach to methylated CpG islands, changing chromatin structure and thus inducing silencing of genes adjacent to CpG islands (Nan et al., 1998). Furthermore, methylation has been directly implicated in phenotypic diversity of hominids. In a recent study, Enard (2004) compared methylation patterns of genes expressed in the brain of *H. sapiens* and *P.troglodytes* and found that more genes were methylated in *H. sapiens* than in *P.troglodytes* brains. These findings complement previous studies conducted by this group, which showed differences in gene expression levels between these latter two species, which suggest a role for methylation in phenotypic evolution.

There are several post-transcriptional mechanisms which can also modify the levels of protein in a cell (Wei et al., 2004). These mechanisms can change the structure and stability of mRNA and can dramatically reduce the levels of mRNA

available for cellular processes. None of these mechanisms were addressed in this thesis, thus they will not be discussed further.

1.8 Evolution of phenotype by changes in *cis* regulatory domains of gene transcription

TFs can bind to specific enhancers/repressor domains, which contain multiple sequence specific DNA binding sites, determining in part the specificity of gene expression. Although the short recognition sites for TF seem vulnerable to random mutation events, the general organization of some promoter elements, including binding sites for TFs within this region and proximal regulatory domains, have been conserved for over 10^7 years (Stone and Wray, 2001). Having said that, turnover of binding sites for TF have been also observed to occur at faster rates than the above suggested time (Dermitzakis and Clark, 2002). For example, comparison of sequences of closely related species and even amongst individuals of the same population shows that binding sites can appear and disappear frequently (Segal et al., 1999). Indeed, comparison of the binding sites found in orthologous regulatory domains tested (activity experimentally verified *in vitro*) of *H. sapiens* and *M. musculs* genes show extensive divergence, with some studies estimating that up to 32%–40% of the *H. sapiens* functional sites are not functional in *M. musculs* (Clark et al., 2003; Dermitzakis and Clark, 2002). This is evidence that there is widespread turnover of transcription factor binding sites (TFBS) at regulatory domains.

Several mechanisms can induce turnover of TFBS. If the binding sites of enhancers of closely related species differed in one binding site, this change is likely to have occurred though local point mutations, whilst when several binding sites differed this is likely to have originated as the result of recombination and retrotransposition (Wray et al., 2003). Even the change of one base or one binding site

for a TF within a regulatory domain could be subject to selective pressure; however, it is noteworthy that TFBS can accumulate substitutions that produce little consequences on an organism phenotype (gene expression) or fitness (Stone and Wray, 2001). This is possible since often, TFs bind to consensus sequences thus can tolerate the occurrence of mutations. Nevertheless, the study of the variation of TFBS constitutes an important tool in the understanding of the molecular evolution of a phenotypic trait.

Examples of the role of changes *cis* regulatory elements in the evolution of gene expression profiles and phenotype are rapidly accumulating in the literature. In modern humans, *cis* regulatory domains of the CCR5 (located in intron 1 and 2), and in the insulin INS (a tandem repeat in its 5' promoter region) genes exhibit signatures of having been positively selected during evolution (Fullerton et al., 2002; Bamshad et al., 2003). This evidence indicates that these *cis* acting domains have contributed to the fitness of modern humans. Genes involved in the regulation of brain function are not an exception for this phenomenon. Positive selective pressures in the *cis* regulators of several genes involved in behavioural regulation, cognition and personality e.g. in the serotonin transporter gene SLC6A4 (Gelernter et al., 1999), dopamine receptor D4 gene DRD4 (Wang et al., 2004b) and prodynorphin PDYN genes (Rockman et al., 2005) have also been identified. Together these studies suggest that *cis* regulatory evolution could have shaped the functional evolution of modern human cognitive capacities.

1.9 Changes in *Trans* regulation as a mechanism for evolution of gene expression

The differential gene expression observed for a particular gene not only reflects changes in *cis* regulation but could also indicate changes in the TFs that interact with them (Wray et al., 2003). Change of *trans* regulators of gene expression

can originate from mutations which affect the expression profile of an upstream TF (whose effect have often greater consequences than when mutations are located in *cis* elements as it affects many genes), typically altering the expression of its promoter targets downstream (Brickman et al., 2001). The change in *trans* regulation can also originate from mutations which take place in the DNA binding domain of an upstream TF thus producing phenotypic change. These events are rare, since there is a higher constrain in conserving DNA binding sites due to the mentioned pleiotropic consequences (Carroll, 2005; Latchman, 1996). Finally, mutations affecting a protein-protein interaction domain in a TF can also affect gene expression; as observed in the human TF FOXP2 (Enard et al., 2002b; Stroud et al., 2006). In modern humans, the FOXP2 shows an evolutionary difference in a phosphorylation site with respect to the FOXP2 of other hominids and this change has been linked to the evolution of language in modern humans.

1.10 *Cis* regulatory elements of gene expression

1.10.1 Non-coding evolutionary conserved regions (ECRs)

Genome wide comparisons of vertebrates have revealed that besides exonic and 5' proximal promoter regions, there are non-coding elements that are greatly conserved across evolution, equally or more so than the coding regions (Davidson et al., 2006; Nie et al., 1996). In this thesis the term ECR(s) will be used to refer exclusively to those located in non-coding or non-proximal promoter regions. The great conservation of these sequences has been shown to be a useful indicator of increased likelihood of gene regulatory activity (Davidson et al., 2006; Pennacchio et al., 2006; Mackenzie and Quinn, 2004). The identification of ECRs has been based on the principle that, because of selective pressure, mutations are expected to accumulate faster in non-functional nucleotide sequences whereas functional sequences (e.g.

promoters, exons or regulatory domains) would remain conserved (Santini et al., 2003). Accordingly, when tested many ECRs have been found to act as regulators of expression of the gene where they are located, and as long-range enhancers, even when located hundreds of kilobases from their target gene (See Figure 1.8a, Davidson et al., 2006; Prabhakar et al., 2006b). Initial studies on ECRs focused on determining the evolution of morphological diversity of vertebrates. In the 1980's studies identified a few gene families (e.g. Homeotic HOM-C and homeobox Hox) important to determine the body plans of all vertebrates, which had been conserved throughout evolution (Carroll, 2000) . However, it was not known how the great morphological diversity found in modern vertebrates would have arisen. In the 1990's, Belting (1998), McGinnis (1990) and others identified *cis* regulatory elements located in the Hox8 gene which had undergone mutation in different vertebrate groups which could affect gene expression during development. Now, several studies have found a successful correlation between variation in ECRs' primary sequence, their gene expression patterns in transgenic animals and change of morphological characteristics (Carroll, 2000). Recent studies suggest a role of ECRs in brain development and function (Pennacchio et al., 2006; Kurokawa et al., 2006). For example, two ECRs found ≈92 and 75Kb upstream the 5'promoter the orthodenticle homeobox gene Otx2, established in the common ancestor of vertebrates. These ECRs appear to be necessary for the correct expression of the Otx2 gene in the anterior neuroectoderm (Kurokawa et al., 2006). The present evidence suggest that mutations in functional ECRs located in genes expressed in the central nervous system (CNS) that may have emerged during the evolution of *H. sapiens* could have contributed to the variation of their patterns of expression, and may correlate with differences from the expression patterns found in other hominid species (Prabhakar et al., 2006b).

Although many human regulatory domains are found conserved between mammalian and vertebrate genomes, some regulatory domains have evolved more recently in the primate genomes, thus would only be distinguishible in comparisons of primate genomes. These type of genomic comparisons also known as "phylogenetic shadowing" (Figure 1.7) has permitted the identification of such significantly divergent functional sequences (Bofelli et al., 2003; Wang et al., 2007). Recently, Wang and collaborators (2007) used this method and successfully identified 7 conserved elements (in 500 kb extracted from several chromosomes). Three of seven of these ECRs (in the low density lipoprotein receptor LDLR; cytochrome P450; family 7, subfamily A, polypeptide 1 CYP7A1 and Sterol regulatory element-binding protein SREBF1 genes) were found to have gene regulatory activity in a variety of *in vitro* and *in vivo* assays. The findings are suggestive that this comparative method can reveal *cis* regulatory domains that have evolved in recent primate history. The potential identification of changes in ECRs present in brain-expressed genes could indicate a recent evolutionary change in their *cis* regulatory properties; perhaps correlated to the cognitive evolution of primates and modern humans. It becomes clear that as observed for the body plans of vertebrates with Hox genes, the evolution of the sequence and regulatory abilities of ECRs, either conserved across vertebrates or across primates could have contributed to the diversification of behavioural phenotype during embryonic development.

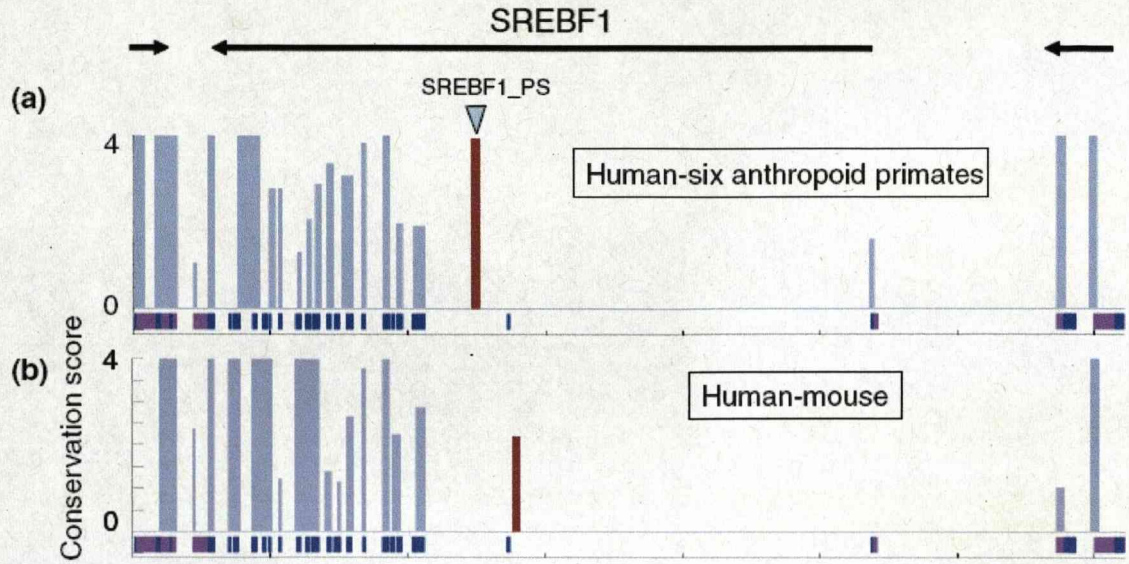


Figure 1.7 Conservation profiles of the SREBF1 gene locus, using comparisons of closely related (modern human-non-human primate) and distantly related (modern human-lab mouse) species. (a) Comparison of the sequences of the SREBF1 genes (with flanking genes partially shown) of seven species spanning 3 major primate groups (modern human, baboon, colobus, marmoset, dusky titi, owl monkey, and squirrel monkey) and (b) Comparison of the sequences of the SREBF1 genes of modern human- and lab mouse. Sequence conservation was calculated using Gumby and visualized using RankVISTA with the modern human sequence as reference. Vertical bars above the horizontal axis depict evolutionarily conserved sequences, with height indicating the conservation score (-log(conservation p value). Coding exons (dark blue) and untranslated regions (UTRs; magenta) are marked below the horizontal axis. Vertical bars that overlap coding exons or UTRs are colored light blue, while non-overlapping bars are colored red. The arrowhead denotes SREBF1_PS, an ECR conserved in primates (p value ≤ 0.005) but not in the mouse (p value > 0.1). Extracted from Wang et al., 2007.

1.10.2 Classes of functional polymorphisms

Mammalian genomes have a great number of polymorphisms equally distributed in coding and non-coding DNA (Conneally, 1994; Stone and Wray, 2001; Wray, 2007). Some classes of polymorphisms are associated with gene regulatory regions, thus are implicated in the variation of gene transcription. *Cis* acting regulatory domains can became polymorphic by a diversity of mutational events, the most common cause being single nucleotide polymorphisms (SNPs) and satellite

DNA (Stone and Wray, 2001). In the modern human genome, SNPs account for 72 % of total variation found (Figure 1.8b). SNPs that occur in TFBS in non-coding regulatory domains have been known to modify the regulatory domain ability to support reporter gene expression. For example, Myers (2007) demonstrated that SNPs in the regulatory region of the human serotonin receptor gene 2 (5-HT2A) might contribute to altered levels of 5-HT2A receptor and to psychiatric disease. However, not in all cases does the change of a nucleotide substitution cause variation of gene expression (e.g. Murgatroyd 2004).

The second most common form of polymorphisms in *cis* regulatory regions of the human genome is satellite DNA (including microsatellite and minisatellite, Figure 1.8b) and represent up to 19-25 % of the total variation found in modern human genomes (Microsatellite consortium webpage: www.microsatellites.org). Within the category of minisatellite DNA there is a subclass, commonly termed variable number of tandem repeat (VNTR) by the clinical literature. As many types of repeats are included under this umbrella term VNTR, in this thesis I will refer to VNTRs following the definition established by Haddley (2007). This is that "the repeats forming VNTRs have sufficient DNA sequence, for example greater than 6 bp, to act as a sequence specific DNA binding site for proteins such as TFs, and therefore have the potential to act as transcriptional regulatory domains".

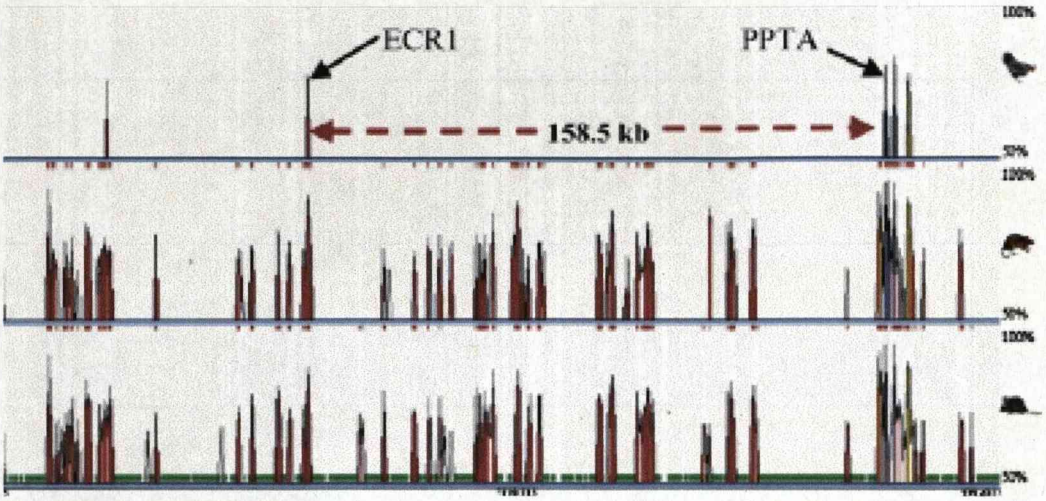
VNTRs are highly variable in repeat number and primary sequence (Berg and Olaisen, 1993; Cerrone et al., 2004; Haddley et al., 2007). This variability is possibly caused by VNTRs replication, which occurs through recombination. Such recombination allows the expansion of VNTRs in length by unequal exchange, either between homologous chromosomes or within chromosomes between sister-chromatids (Harding et al., 1992). Recombination between non-identical VNTR

homologous alleles can occur, generating both sequence conversion and new copy number allelic variation (Jeffreys, 1997). The new alleles often differ in number and type of TFBS. This variation correlates with differential affinities for binding of TF *in vitro* (Heils et al., 1995; Klenova et al., 2004; Lovejoy et al., 2003) and differential reporter gene expression *in vitro* and *in vivo* (Battersby et al., 1996; D'Souza et al., 2004; Guindalini et al., 2006; Heils et al., 1997; Pugliese et al., 1997; Reif and Lesch, 2003).

VNTRs have been found in introns, in the proximity of promoters, in exons and in unstranslated regions of genes associated with human and non-human primate behaviour. Tere are examples where VNTR variability correlates with variation of behaviour within modern humans and primate populations (Barr et al., 2003; Hariri and Holmes, 2006; Miller et al., 2004; Miller et al., 2001; Shimada et al., 2004; Wendland et al., 2006a; Wendland et al., 2006b). Since behaviour and cognition have high adaptive value, it is possible these VNTRs are implicated in the evolution of behavioural responses in primates. It has been demonstrated that intronic and promoter VNTRs can play a role in gene transcription (e.g. Mackenzie and Quinn 1999; Meloni et al., 1998). Additionally, VNTR can also participate in regulation of gene expression at other levels. For example, VNTRs located in UTR regions have also been suggested to play an important role in mRNA stability, which can also affect levels of transcript available in the cellular environment (Nakamura et al., 1998; Zamorano et al., 2006). Further, VNTRs located in exons are known to affect protein folding and have been associated to differential levels of mature RNA availability in the cytoplasm (Oak et al., 2000).

In following sections, I will discuss the evidence, which suggested a role for VNTRs as active contributing factors in the diversification of behavioural phenotype and cognition via modulation of gene expression in humans and non-human primates.

a



b

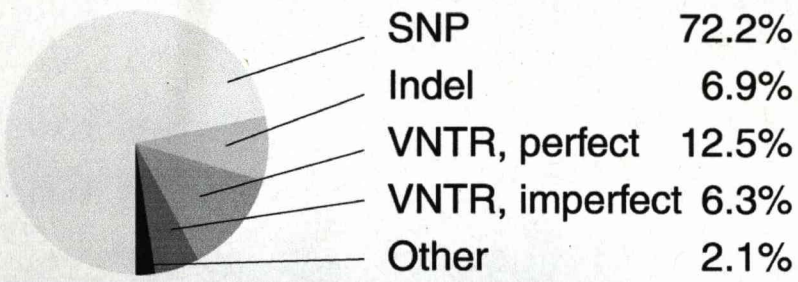


Figure 1.8 Types of *cis* regulators of gene expression. (a) Example of an ECR. An ECR 158 kb upstream the preprotachykinin (PPTA) gene (extracted from Davidson et al., 2006) was identified using the ECR browser **(b) Polymorphisms in *cis* regulatory regions found in the human genome.** SNPs cause greatest variation at regulatory domains followed by VNTRs (Extracted from Rockman and Wray, 2002).

1.10.3. VNTRs in neurotransmitter genes

The genes involved in the regulation of levels of dopamine (DA) and serotonin (5HT) in modern human and non-human primates present allelic variation. This variation is located in their coding regions (e.g. DRD2, DRD4 and HTR2 genes) and more often; this is located in non-coding regions (e.g. SLC6A4, SLC6A3 DRD4 and MAOA genes). The allelic variability conferred by these VNTRs and SNPs in the latter genes have been shown to correlate with differential gene expression *in vivo* (Hariri et al., 2002a; Hranilovic et al., 2004) and when tested they can support differential reporter gene expression *in vitro* and *in vivo* (Inoue-Murayama et al., 2002a; MacKenzie and Quinn, 1999; Roberts et al., 2007). This variability has also been associated with variation of dopamine and serotonin neurotransmission and related behaviour in humans and non-human primates (Hranilovic et al., 2004; Mill et al., 2002). In the following sections, I reviewed the literature for VNTRs located in neurotransmitter genes of human and non-human primates associated with cognition and behaviour.

1.10.3.1 VNTR in the SLC6A4 gene

The serotonin transporter protein (5HTT) is a sodium activated membrane transporter involved in active reuptake of serotonin in the post-synaptic space, which is a target for several antidepressants, amphetamines, and potent neurotoxins. Efforts have focused on the role of the SLC6A4 gene on psychiatric disorders as it encodes for the 5HTT protein. The organization of this gene was determined by Lesch and colleagues (1994), where they presented data on the coding and promoter sequence of the human SLC6A4 gene. It was shown that this is gene is approximately 38 kb long, is formed by 14 exons, and presents many polymorphisms (Figure 1.9). This study also found a VNTR 1.2 kb upstream the 5' promoter, composed by 20-23 bp/repeat

units (Figure 1.9). The most commonly variants of this VNTR present 14 or 16 repeats, but SNPs and variants with 15, 18, 19, 20 and 21 repeats units have been identified (Sakai et al., 2002). It has been demonstrated that these VNTRs were capable of acting as *cis* regulators of reporter gene expression (Heils et al., 1996) and several studies have since confirmed the correlation between the 2 common alleles and diversity of human behaviour (Curran et al., 2005; Hariri et al., 2002a; Mulder et al., 2005). To validate the role of *cis* regulatory domains in the mediation of a specific behavioural trait, scientists have usually used animal models. However, the absence of this VNTR from the SLC6A4 gene from the lab mouse and rat has resulted in the use of *Macaca mulatta* (rhesus macaque) which exhibits a homologous VNTR. Thus, this species has been used for the conduction of epidemiological and pharmacological studies (Barr et al., 2003; Barr et al., 2004), which analyse the role of this VNTR, and serotonin related behaviour. Similarly as found in modern humans, the 2 alleles of the *M. mulatta* promoter VNTR support differential transcriptional activity *in vitro* (Bennett et al., 2002) which has been correlated with variation of emotional behaviour *in vivo* (Barr et al., 2004; Bennett et al., 2002).

Studies have shown that this VNTR is also present in the 5' promoter region of the SLC6A4 gene of great apes (family Hominidae), lesser apes or gibbons (family Hylobatidae) and old world monkeys (Family Cercopithecidae), but not in primates more distantly related to modern humans (Lesch et al., 1997). It has been proposed that these VNTR could have evolved in the promoter region of the SLC6A4 gene of primates to produce a greater level of complexity in the serotonin related behaviour in this highly evolved and socially complex group of mammals. Furthermore, Lesch and colleagues (1997) have speculated that this VNTR could have originated from

retrotransposition of mobile element or by insertion of viral particles in primate genomes before the ancestors of old world monkeys, apes and humans separated.

In the second intron of the SLC6A4 gene of modern humans a second VNTR has been identified (Lesch et al., 1994, Figure 1.9). This VNTR presents commonly 9, 10 or 12 repeat units, composed by 16/17 bp/repeat, although VNTRs with 7 and 11 repeat units have been reported (Soeby et al., 2005; Gelernter et al., 1999). Similar to the 5' promoter VNTR, the 3 commonly found alleles of the STin2 VNTR are capable of supporting differential transcriptional activity *in vitro* (Haddley et al., 2007; Klenova et al., 2004; Lovejoy et al., 2003; Roberts et al., 2007). Furthermore, the capacity of the 10 and 12 repeat VNTRs to support differential transcriptional activity *in vivo* was demonstrated using a transgenic mouse embryo system (McKenzie and Quinn 1999). Interestingly, these constructs were active at the initial site (rostral midbrain) of serotonergic differentiation in the mouse. The correlation between the STin2 VNTRs genotype and predisposition to develop neurological disorders or to present extremes of normal behaviour has been investigated in modern humans. There are many studies suggesting that this STin2 VNTR may play a role in the predisposition to stress, anxiety, addictive behaviour, aggression and impulsivity and cognitive decline (Kremer et al., 2005; Mulder et al., 2005; Payton et al., 2005). However, these correlations have sometimes been disputed (Lasky-Su et al., 2005). As it will be discussed later on this thesis, the discrepancies in results may rise as more than one polymorphic regulatory domain is likely to be involved in the transcriptional regulation of the SLC6A4 gene and therefore, the study of multiple domains should be addressed.

Soeby and colleagues (2005) investigated the presence of this STin2 VNTR in other mammals. This study showed that this VNTR was present only in other great

apes (presenting from 19 to 40 repeats) and old world monkeys (exhibiting 5 repeats) but not in primates more distantly related to humans or rodents. Furthermore, this group conducted an *in silico* analysis of the primate STin2 VNTR sequences and suggested that the different number of TFBS occurring in the different species might correlate with different transcriptional activity of these VNTRs. Consequently, it is possible that the differences in the primate STin2 VNTR could have consequences in the regulation of the SLC6A4 gene expression *in vivo* (Soeby et al., 2005).

The effects of the genotype of these two polymorphic loci in the SLC6A4 gene are correlated with variation of normal behaviour in human and primate populations. For example, Trefilov (2000) has shown that in *M. mulatta* (rhesus macaques) there is a VNTR with two alleles termed long (l) and short (s). Interestingly, the genotype of male macaques appears to correlate with reproductive success as males heterozygous (l/s) for the promoter VNTR sire more offspring that those homozygous for either the small (s/s) or long allele (l/l). This has been explained as that, individuals carrying s/s tend to leave their natal group earlier than those carrying s/l or l/l. Consequently, s/s individuals are exposed to a higher mortality risk before mating age in comparison to s/l or l/l individuals. However, those carrying s/s who survived are often preferred by female *M. mulatta* for mating. Similarly, individuals carrying the l/l genotype tend to remain in their natal group for longer than s/s and s/l individuals, and although have higher survival rate before mating age, are selected against for females seeking a mating partner. The authors proposed that as neither s/s nor l/l are optimum 5' promoter genotypes for *M. mulatta*, therefore it is the l/s the genotype that is favoured by natural selection.

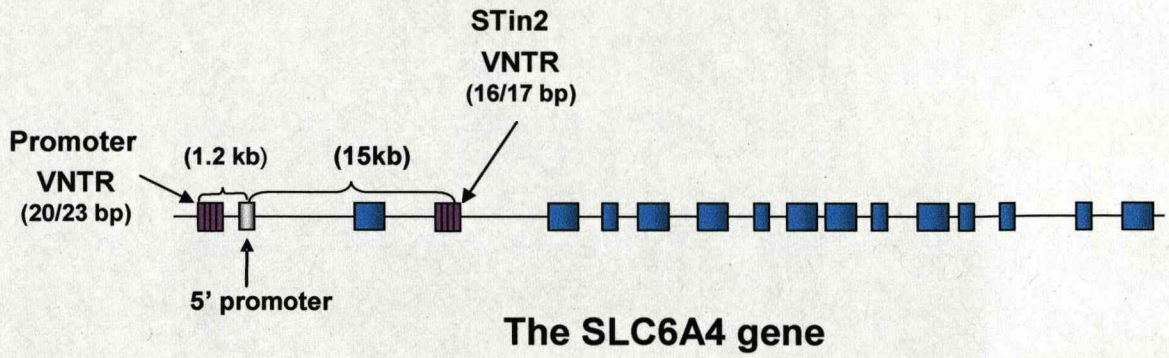


Figure 1.9 The SLC6A4 gene and its VNTRs. The blocks in blue represent exons, the blocks in fuchsia represent VNTRs. The core 5' promoter of this gene (represented as a block in grey) has been identified 1.2kb from the initiation start site (extracted from Lesch et al., 1994).

Due to the influence of society on modern human behaviour, studies on the influences of genes on human behaviour are intrinsically more complex in humans than in other animals. Interestingly, the correlation between genotype and behavioural phenotype become evident in individuals that have suffered stressful life events. For example a few studies have suggested a correlation between the short allele of the 5' promoter VNTR and increased aggression, increase tendency to exhibit risky behaviour e.g. use of amphetamines, risky sexual behaviour, gambling (Galeeva et al., 2002); sport performance (Park et al., 2004) and shyness in children (Arbelle et al., 2003). In addition, the presence of this allele has been identified as a risk factor for depression (Caspi et al., 2003; Eley et al., 2004). The 10 and 12 repeat alleles of the STin2 VNTR have been alternatively linked to attention deficit (Kim et al., 2005), creativity (Bachner-Melman et al., 2005) and anxiety (Ohara et al., 1999).

1.10.3.2 VNTRs in the DRD4 gene

The dopamine receptor D4 is a G protein coupled receptor member of the D2-like dopamine receptor family, that plays a crucial role in mediating the diverse

effects of dopamine in the CNS (Van Tol, 1998). In 1992, Van Tol described the organization of this gene and showed that it presents four exons and contains a number of polymorphisms in its coding and non-coding regions. The most studied polymorphism in this locus is a VNTR composed by 2 to 11 repeat units (48 bp/ repeats or 16 amino acids, Figure 1.10) is present in the third exon. This VNTR affect the length and sequence of the third cytoplasmatic loop of the receptor. The 3 most common variants of this receptor (with 2, 4 or 7 repeats per VNTR) exhibit different functional characteristics. For example when analysed *in vitro* in CHO cells, the potency of dopamine to inhibit cyclic AMP (cAMP) formation was about two fold reduced between the variants with the 7 repeat VNTR compared with those with 4 and 2 repeats (Asghari et al., 1995). The transcriptional activities of the different alleles of this VNTR were later demonstrated by Schoots and Van Tol (2003). They showed that these 3 most common VNTR variants were capable of supporting differential reporter gene expression in GH4C1 cells (rat pituitary) which express endogenous DRD4 gene (Schoots and Van Tol, 2003).

The presence of this VNTR in the DRD4 genes of non-human primates was investigated by Livak (1995). This study showed that this VNTR was polymorphic in each species of great ape, old and new world monkey analysed. Later, Inoue-Muruyama (1998) investigated the presence of this VNTR in more distantly related primates (prosimians and tarsiers) and in insectivores (tree shrews), a group of animals which are the earliest relatives to the entire order primates. Inoue-Muruyama (1998) found that the VNTR was present in prosimians; however, the 48 bp sequence was not replicated in tree shrews DRD4 gene. This demonstrated that this sequence only began to expand into a VNTR ~60 mya, with the origin of primates. Analogous repeats to the primate VNTR have been identified in the DRD4 genes of other

mammalian groups such as cetaceans, carnivores and artiodactyls. Based on these findings it has been proposed that this locus is a hotspot for recombination and is associated with the diversification of dopamine related behaviour in mammals (Larsen et al., 2005; Mogensen et al., 2006; Momozawa et al., 2005).

More recent studies have identified other VNTRs in the DRD4 gene of humans and non-human primates (Seaman et al., 2000; Seaman et al., 1999; Shimada et al., 2004). The first is a VNTR located 1.2 kb further upstream of the major translational start site (Figure 1.10). This has evolved from a duplication of 120 bp which only occurring within modern human populations, and in this duplication is absent from great apes or monkeys (Seaman et al., 1999). Gene expression assays have shown that the two commonly found alleles of the 5' promoter VNTR are capable of supporting differential reporter gene expression (D'Souza et al., 2004) as seen for the VNTR in the exon 3.

The DRD4 gene presents a second VNTR located in a coding region. This is composed by 12 bp/repeat (1 and 2 copies or the repeat) and is found in the first exon. This VNTR is also present in great apes and old world monkeys but is absent from the DRD4 gene of new world monkeys (Seaman et al., 2000). The most recently discovered VNTR in this gene was identified in the 2nd intron and involves the expansion of short GC rich 6 bp long sequence. This VNTR has only been studied in a small number of specific human populations (Japanese and Hungarian) and it has been demonstrated that its presence is variable amongst these two groups (Shimada et al., 2004). Polymorphic regions homologous to this elements in *H. sapiens* have also been identified in other primates. In spite of the potential importance of the latter two VNTRs in regulating the DRD4 gene expression, its functional activities are yet to be explored.

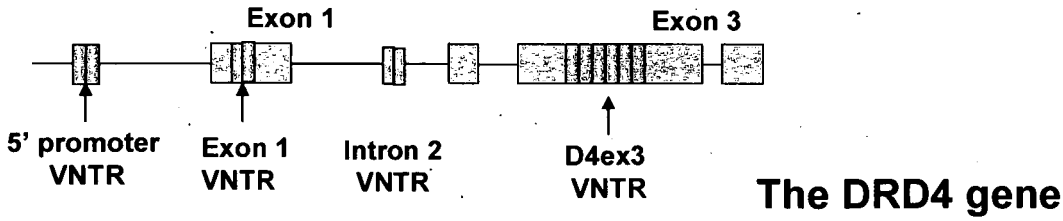


Figure 1.10 The DRD4 gene and its VNTRs. Blue blocks represent exons and fuchsia blocks represent VNTRs identified in this gene. The gene does no contain a canonical TATA box, but a negative modulator has been determined in the 5' flanking region (1.2 kb) from the initiation codon (modified from Seaman et al., 1999 and Kamakura et al., 1997).

Due to the effect of the exon 1 and 3 VNTRs on the length and structure of the DRD4 protein and the potential *cis* regulatory properties exhibited by 5' promoter and exon 3 VNTRs, many association studies have tested possible correlations between specific copy number variants of these VNTRs and dopamine related behaviour in primates. These studies have suggested a correlation between traits such as: novelty seeking, attention deficit and even historical events such as the distance migrated by ancient human populations and these VNTRs; however, these results have often been disputed (Aguirre et al., 2007; Bailey et al., 2007; Congdon et al., 2007; Lakatos et al., 2002; McCracken et al., 2000). The SNPs present in these VNTRs may affect the function of the copy number variants of these VNTRs, thus affecting the results of association studies. Although it is possible that the correlation between behaviour and the exonic polymorphisms of the DRD4 gene emerge from the effect of the VNTR on the protein structure and function, these could also originate from the differential gene expression capacities exhibited by these VNTRs.

1.10.3.3 VNTR in the SLC6A3 gene

Dopamine (DA) normal homeostasis in the brain is regulated in part by the uptake of DA via the dopamine transporter protein (DAT). DAT is a plasma

membrane protein, which is encoded by the SLC6A3 gene and belongs to the Na+/Cl−dependent family of neurotransmitter transporters and is the primary target for cocaine and amphetamine and for therapeutic agents used in the treatment of neurological disorders, such as attention deficit hyperactivity disorder (ADHD) and depression. The SLC6A3 gene is organised in 15 exons, separated by 14 introns; presents several VNTRs in introns, untranslated regions (UTRs) and SNPs in both intron and exon sequences (Greenwood and Kelsoe, 2003; Guindalini et al., 2006; Vandenbergh et al., 1992).

Many studies addressing genetic predisposition to substance addiction, have mainly focused on the 3' UTR VNTR, first described by Vandenbergh in 1992. This VNTR can be composed of 3 to 14 repeat units formed by 40 bp/repeat, the most common variants being those with 9 and 10 repeats (Greenwood and Kelsoe, 2003). Several association studies have shown that the 9 or 10 repeat VNTR are risk factors for the development of addiction to cocaine, amphetamines and alcohol, ADHD and deficiencies of motor and cognitive abilities (Cook et al., 1995; Gelernter et al., 1994; Sander et al., 1997). Study of this 3'UTR VNTR in non-human primates suggests that this domain is primate specific, as this VNTR have been found to be polymorphic in all apes, old and new world monkeys studied (Inoue-Murayama et al., 2002a, Miller et al., 2001) but this is absent from the rodent SLC6A3 genes (Miller et al., 2001).

The transcriptional activities of the 3'UTR VNTRs have been assessed *in vitro* in dopaminergic cell lines such as the SN4741, and SK-N-SH (Inoue-Murayama et al., 2002a; Greenwood and Kelsoe 2003; Miller and Madras 2002 and Michelhaugh et al., 2001) and in organotypic cultures of rat midbrain (Michelhaugh et al., 2001). These studies have demonstrated that the 9 and 10 repeat VNTRs are capable of supporting differential levels of reporter gene expression in DAT expressing cells

when cloned in vectors containing a viral promoter (e.g. thymidine kinase TK or simian virus 40 SV40 early promoters). However, a study that analysed the transcriptional activity of the same 3' UTR VNTR variants cloned into a luciferase vector driven by the SLC6A3 promoter into HEK293 cells, was unable to replicate the previous findings (Greenwood and Kelsoe 2003). It is worth noting that in this latter study the DAT 9 and 10 constructs included flanking regions of considerable size, which may have harboured repressor elements affecting VNTR function. These results provoked a debate on the role of this VNTR in SLC6A3 expression; however, several association studies, ligand binding assays and *in vivo* studies support the role of the 3' UTR VNTR in creating significant differences in the expression of the SLC6A3 gene (Mill et al., 2002; Heinz et al., 2000; VanNess et al., 2005). Furthermore, Inoue-Murayama and colleagues (2003) showed that constructs containing VNTRs of great apes and old world monkeys tested supported significantly different levels of gene expression (with respect to a control) and interestingly, the constructs containing the *H. sapiens* and *P. troglodytes* VNTRs showed marked differences in their transcriptional activities *in vitro*. These findings would indicate that these VNTRs could correlate with species-specific levels of reporter gene expression, and contribute to create differences in SLC6A3 expression amongst these species. However, these VNTRs present abundant SNPs, which can affect their transcriptional activity of *in vitro* (Miller et al., 2001) that also need to be addressed.

A second VNTR in the intron 8 of the human SLC6A3 gene has also been correlated with addictive personality in a Brazilian sample. Furthermore, different alleles of this VNTR demonstrate differential reporter gene activity (Guindalini et al,

2006). The presence of this VNTR in mouse or non-human primate genes has not been explored to this date.

1.11 TFs relevant for the regulation of VNTR in neurotransmitter genes

Many studies have attempted to demonstrate the interaction between VNTR binding motifs and TFs. For example, our group (Michellaugh et al., 2001) demonstrated that the STin2 VNTR and the 3' UTR VNTR of the SLC6A4 and SLC6A3 genes (respectively) present similarities in their sequence and have related protein binding properties. Oligonucleotides representative of either VNTR repeat sequence cross-compete each other for related protein complexes. This suggested that although these VNTRs have distinct binding properties, they could regulate transcription via interactions in part with the same regulatory proteins. The interaction between binding sites within the STin2 VNTR and specific TFs has been demonstrated *in vitro* and *in vivo* (Roberts et al., 2007; Klenova et al., 2004).

A brief description of these TFs and the findings on their interactions with neurotransmitter VNTRs are detailed in the following sections.

1.11.1 Y-box biding protein YB-1

YB-1 is a ubiquitous multifunctional transcriptional factor, which belongs to the cold shock-domain protein superfamily (CSD). YB-1 is pluripotent protein, developmentally regulated TF and believed to participate in transcriptional and translational regulation, DNA repair and drug resistance and be involved in the signal transduction pathway for environmental stress and drug resistance (Kohno et al., 2003). For example, YB-1 translocates from cytoplasm to nuclei in the presence of stress stimuli and is involved in mediating effects of stress responses to ultraviolet (UV) irradiation and interleukin-2 treatments (Kohno et al., 2003; Uchiumi et al., 1993). In its role as a TF, YB-1 has been identified as an activator of promoter

function of a large variety of genes (Mertens et al., 1998) but can also act as a repressor of gene expression e.g. of the human multidrug resistance 1 (MDR1) and of the α-2 (I) collagen genes (Norman et al., 2001). YB-1 typical binding sites are CCAAT boxes and additionally, YB-1 can directly regulate gene expression via interaction with Y-boxes (CTGCTGGGCAAG; Kohno et al., 2003; Klenova et al., 2004; Figure 1.11a) or though interactions with other TFs (e.g. YY1, AP2, CTCF, PAX6 Figure 11.a and 11.b). Such interactions can induce differential modulation of the transcription of a variety of genes. For example, AP2, a differentiation-dependent TF, can directly interact with YB-1 to induce a synergistic activation via a response element in the 5' region of the rat gelatinase A gene (Mertens et al., 1998). Although the binding of YB-1 has high specificity, it has been shown to bind a range of binding sites or motifs in cellular and viral genes (Sawaya et al., 1998).

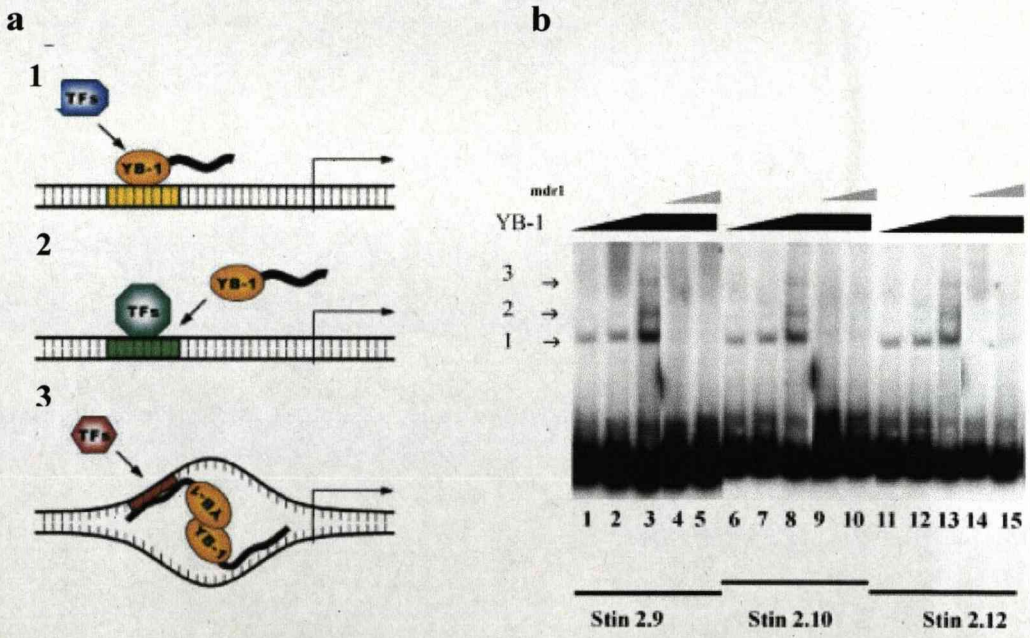


Figure 1.11 Transcriptional control by YB-1 (a) 1. YB-1 directly binds to Y-box (in yellow) and related sequences or directly binds to Y-box with other TFs. 2: YB-1 interacts with other TFs and functions as either co-activator or co-repressor. 3: YB- 1 binds to the single-stranded region of the promoter either to enhance or inhibit the DNA binding of TFs (extracted from Kohno et al., 2003). **(b)** Analysis of the YB-1 interactions with the STin2VNTR domains in EMSA. Bacterially expressed YB-1 can form specific complexes with the 32P-labeled Stin 2.9, Stin 2.10, and Stin 2.12 DNA (From Klenova et al., 2004)

1.11.2 CTCC binding factor CTCF

CTCF is an 11 zinc finger (ZF) protein, which is highly conserved from drosophila to man (Ohlsson et al., 2001). CTCF is ubiquitously expressed in all somatic cells of vertebrates and is able to bind to varying target sequences to perform different regulatory roles. This includes transcriptional activation or repression of promoters, hormone-inducible gene silencing, and creation of constitutive or methylation-sensitive chromatin boundaries (Bell et al., 1999). CTCF has also been found to act as an insulator, generating functional block to enhancer-promoter interactions (Bell et al., 1999) and also it has been suggested to act as a tumor-suppressor gene (Filippova et al., 1998).

The repressor/activator role of CTCF in transcription is dependent on its binding within the context of the promoter. For example, in corneal epithelial cells, CTCF has been observed to repress transcriptional activity of the mouse PAX6 gene (Pair box 6, a developmental transcription factor), presumably through elements 1.2 kb from the 5' promoter (Wu et al., 2006). Conversely, CTCF has been seen to act as an activator of expression of the amyloid beta-protein precursor promoter (Vostrov and Quitschke, 1997). The effect of CTCF on transcription is also modulated by the thyroid hormone receptor (T3R) (Ohlsson et al., 2001). CTCF present silencing domains that mediate repression, and it is capable of interacting with other factors that modify histone deacetylase activity, altering chromatin structure and therefore, plays a major role in epigenetic changes of chromatin.

CTCF uses different sets of ZFs to recognize a large variety of different DNA sequences; further, each of the diverse DNA CTCF complexes might associate with different partners to define distinct functions in the cell. Indeed, Chernukhin (2000) demonstrated that CTCF acts with YB-1 *in vitro* and *in vivo* and that this specific

association requires the zinc finger domain of CTCF, resulting in CTCF specific repression of the activity of the c-*myc* reporter activty. The complex formed by these two proteins has been identified in several cell lines. The interaction of these two proteins in the context of VNTR regulation is discussed below.

1.11.2.1 CTCF and YB-1 regulate STin2 VNTR function

Our group investigated the potential TFs that are responsible for the regulation of the STin2 VNTR transcriptional activities using yeast-one hybrid screen. This showed that the Y-box binding protein 1 (YB-1) interacts with this VNTR (Klenova et al., 2004). Furthermore, it was shown by electrophoretic mobility shift (EMSA) and chromatin immunoprecipitation (ChIP) assays that this VNTR can bind to YB-1 *in vitro* and *in vivo*, possibly through two distinct types of Y-boxes ("a" and "g") found within the VNTR variants. Since the number of these Y-boxes varies in each allele, the affinity of the VNTR for YB-1 is allele dependent (Klenova et al., 2004). Given that CTCF acts as a binding partner for YB-1, the authors tested whether CTCF overexpression would affect the regulation exerted on the VNTRs by YB-1. They showed that CTCF abrogated the activator effect induced by YB-1 in co-transfection experiments, and the binding between YB-1 and the VNTR. In a subsequent publication (Roberts et al., 2007), our group showed that differences in binding sites for CTCF found in the STin2 VNTRs have been linked to the differential activation of the levels of reporter gene expression supported by the STin2 VNTR variants induced by over-expression of CTCF *in vitro* in JAr cells. Furthermore, lithium chloride (LiCl), a therapeutic agent for treatment of bipolar affective disorders, affected the CTCF and YB-1 regulation of the STin2 VNTR. CTCF and YB-1 showed both differential binding to the polymorphic alleles of the VNTR and the levels of CTCF, YB-1 and SLC6A4 mRNA and protein were altered *in vivo* in response to lithium

chloride. These results indicate that both YB-1 and CTCF may contribute to regulation of SLC6A4 gene expression via the STin2 VNTR, and pose the possibility of a regulatory role of these two proteins on other VNTRs in neurotransmitter genes which might show related consensus sequences.

1.11.3 Sp1 mediates neuronal genes expression

Specificity factor 1, Sp1, is the founder member of the Sp family of TFs. As all members of the Sp family, Sp1 possess 3 zinc fingers which act as its functional domains for binding DNA and shows great conservation across vertebrates (Suske et al., 1999). Sp1 plays a crucial role in the formation of initiation complex for transcription (interacting with TAFs) of many viral and cellular genes and in transcription, via binding to GC boxes often located in proximity to the TATA box (Bouwman and Philipsen, 2002). Binding sites for Sp1 close to the initiator elements have also been shown to enhance the levels of transcription (Seto et al., 1993). This activation can also occur though enhancers located outside the promoter region (Cohen et al., 1997; Talianidis et al., 1995). The transcription activation induced by Sp1 is aided by the formation of protein-protein interactions with a variety of TFs (e.g. YY1, PAX6) and it has been demonstrated that such interactions have a synergistic effect on transcription initiation (Seto et al., 1993).

Sp1 has been shown to regulate a great variety of genes expressed in the central nervous system (CNS). Amongst them is SLC6A3 (Wang and Bannon, 2005), the neuronal nicotinic receptor subunit β4 (Bigger et al., 1997), NMDA receptor subunit type 1 (Okamoto et al., 2002), serotonin 1A receptor (Parks and Shenk, 1996) and the nitric oxide synthase NOS I genes (Bachir et al., 2003). Interestingly, Sp1 activates transcription of several viral promoters via interactions with repeated sequences similar to primate VNTRs found in neurotransmitter genes (Barnhart et al.,

1997; Chen-Park et al., 2002; Suñé and García-Blanco, 1995). This has encouraged several *in silico* and *in vitro* studies, which investigated the presence of Sp1 sites in neurotransmitter gene VNTRs. These have suggested that indeed, there are potential interactions between Sp1 and binding sites within the VNTRs of the DRD4 gene (Ronai et al., 2004; Schoots and Van Tol 2003; Seaman et al., 1999) and the MAOA gene (Inoue-Murayama et al., 2006) which may contribute to the differential *cis* regulatory role of these VNTRs.

1.12 Rationale for the selection of genes and regulatory domains studied in this project

This thesis primarily focused on the analysis of potential *cis* regulators (VNTRs and ECRs) in the expression of the SLC6A4 and DRD4 genes. These genes were selected as previously discussed as their gene products have been linked to cognitive and high executive function of the brain. Additionlly, the behavioural outcome that is regulated by the products of these gene has been reported to differ between *H. sapiens* and the great apes (Dainton and Macho, 1999; Manson and Wrangham, 1991; Stanford, 2006; Varki and Altheide, 2005).

The VNTRs primarily studied in this project were the STin2 and 5' promoter VNTRs of the SLC6A4 gene and the VNTR in the third exon (D4ex3) of the DRD4 gene. These VNTRs were selected because they present similarity in their primary DNA sequences. This type of VNTR with high GC rich sequences have been found associated to diversification of neuronal gene function in a number of genes (Hariri and Holmes, 2006; Heinz et al., 2000). As such, this type of sequences may be a mechanism that has evolved to contribute to the plasticity of gene expression. Potential changes in neuronal function undergone by the human brain during evolution might have been involved this type of sequences.

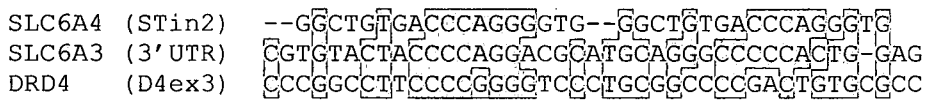


Figure 1.12 Alignment of the sequences of the STin2 VNTR of the SLC6A4 gene, the D4ex3 VNTR of the DRD4 gene and the 3'UTR VNTR of the SLC6A3 gene. The base pairs that were present in more than one sequence were highlighted in yellow. Spacer hyphens were introduced to optimise the alignment. Underlined in the alignment is one of the motifs, which presents potential TF binding properties

1.13 Aims and objectives of this study

The evolution of the cognition of modern humans has been mediated by many different mechanisms of molecular evolution (section 1.4). To date the medical literature has established the *cis* regulatory role of VNTRs and ECRs (e.g. in DRD4, SLC6A4, SLC6A3 and PPTA gene) and have shown their correlation with modern human behaviour (e.g. Guindalini et al., 2006; Lesch et al., 1997). Therefore, the primary aim of this thesis was to investigate how the function of specific *cis* acting regulatory domains such as VNTRs and ECRs found in genes linked to cognitive behaviour SLC6A4 and DRD4 genes have evolved in hominids. First, I aim to investigate if these sequences have undergone change, exclusive to the *Homo* lineage. This was achieved by the comparison of the potential TFBS found in the regulatory domain sequences of modern human and non-human primates. For this purpose, I reconstructed the evolution of the sequences based on the TFBS found within the sequences using a phylogenetic inference software package.

The second aim was to determine if these elements can act as *cis* regulators of gene expression *in vitro* and to compare if the differences or similarities in their sequences that have appeared during evolution correlate with their functional profile *in vitro* in cells derived from the CNS. This was achieved by the cloning of the potential regulatory domains in luciferase reporter cassettes systems and delivered into cultured cells derived from the CNS to analyse their ability to support reporter gene expression.

Chapter 2 Materials and Methods

2.1 Materials

2.1.1 Reagents and Solutions

LB (Luria-Bertani broth)

Tryptone 10 g, yeast extract 5 g, NaCl 5 g and 1 l of H_2O

S.O.C medium

20 g Bacto Tryptone, 5 g Bacto Yeast Extract, 2 ml of 5 M NaCl, 2.5 ml of 1 M KCl, 10 ml of 1 M MgCl2, 10 ml of 1 M MgSO4, 20 ml of 1 M glucose, 1 l of distilled H_2O

LB AGAR

L-B broth with 1% agar

TBE buffer X10

Tris base 108 g, boric acid 55 g, 0.5 M EDTA 20 ml, water to 1 l (pH 8).

6x agarose gel loading buffer

Bromophenol blue 0.25% (w/v), xylene cyanol, 0.25% (w/v), 1 mM EDTA, 30%, glycerol and 70% water

Chelex solution

Chelex ®100 (5% v/v, Bio-Rad) solution, 0.039 M DTT 50μg of Proteinase K and water to 250 μl

Sodium acetate buffer (3M pH 5.2)

408.3 g of sodium acetate-$3H_2O$ in 800 ml of H_2O. Adjust the pH to 5.2 with glacial acetic acid. Adjust the volume to 1 l with H_2O.

2.1.2 Plasmids used during this project

The following table (Table 2.1) contains plasmids (TF expression vectors, VNTR constructs and commercial plasmids) not generated by the author and used during the project.

pGL3p pGL3c	This plasmid is a firefly luciferase expression vector carrying a minimal SV40 promoter. VNTR fragments were cloned into pGL3p to asses their gene expression regulator capacity. The pGL3c plasmid has the same backbone as pGL3p and posses a SV40 enhancer. This was used as positive control of transfections. Both vectors contain an ampicillin resistance gene for selection of transformants in selective medium.
pGEM-t	Vector used for direct cloning of PCR fragments. The vector carried T7 and SP6 promoter and a multiple cloning site, which interrupts the coding sequence of the beta galactosidase gene. pGEM-t (Promega, UK) carries an ampicillin resistance gene for selection of positive transformants in selective medium.
pmLuc2	Renilla luciferase expression vector carrying a minimal TK promoter used as internal control for transfections. It carries an ampicillin resistance gene for selection of positive transformants in selective medium (Novagen).
Stin2.9 luc STin2.10 luc STin2.12 luc	Luciferase expression vectors (pGL3p backbone) carrying the 3 different allelic variants (9, 10 and 12) of the *H. sapiens* SCL6A4 gene intron 2 VNTR (Fiskerstrand et al., 1999).
DAT9.luc DAT10.luc	Luciferase expression vectors (pGL3p backbone) carrying the 2 different allelic variants (9 and 10 repeats per VNTR) of the 3'UTR of the *H. sapiens* SCL6A3 gene (Michellaugh et al., 2001)
Homo14.luc Homo 16.luc	Luciferase expression vectors (pGL3p backbone) carrying 2 different allelic variants (Homo16, with 16 repeat units and Homo14, with 14repeat units) of the promoter VNTR of the human SCL4A6 gene. Generated by Mr. F. Ali and Mr. J. Roberts respectively
CTCF YB-1	Expression vectors with the human full-length cDNA YB-1 and CTCF described previously (Klenova et al., 2004).

2.1.3 Culture media

2.1.3.1 Cell line media

2.1.3.1.1 JAr cell culture medium: JAr, a human choriocarcinoma cell line, was cultured in RMPI 1640 (RPMI 1640 with L-Glutamine Hyclone; Cat. No SH30027.01) supplemented with 10% (v/v) heat-inactivated calf serum (FCS; Autogen Bioclear Cat No 7.01), penicillin/streptomycin (100 units/100 μg) at 37°C in 5% CO_2 humidified atmosphere.

2.1.3.1.2 SN4741 cell culture medium: SN4741, a mouse embryonic substantia nigra-derived cell line, was grown in Dulbecco's modified Eagle's high glucose medium (Bioclear) containing 10% heat-inactivated FCS (Autogen Bioclear), L-glutamine (2 mM) and penicillin/ streptomycin (100 units/100 μg) at 33 °C in 5% CO_2 humidified atmosphere. For both cells lines, the medium was changed every 2 days, and cells were sub cultured about twice a week.

2.1.3.2 Media for culture of primary rat cells

2.1.3.2.1 Culture medium I: DMEM (Bioclear Cat. NoAB2052) containing 10% FCS.

2.1.3.2.2 Culture medium II: Neurobasal-A medium [Invitrogen/Gibco; Cat. No. 10888-022] 2% B27 supplement, 2 mM GlutaMAX I and 500 μg of gentamycin.

2.1.3.2.3 Dissecting solution: 91 ml Hanks balanced salt solution (HBSS. Invitrogen-Gibco BRL cat No 24020-091) containing 3.5 ml 1M HEPES, 1 ml 1 M $MgCl_2$, 1ml 200 mM L-glutamine, 1 ml 100x penicillin/streptomycin (equates to a working concentration of 10,000 units penicillin and 100 μg/ml of streptomycin, Sigma-Aldrich Ltd. cat No P0781).

2.1.4 Animals

Male Wistar albino rats (2-7 days old) were used to generate primary cell cultures from cortical tissue. All animals were purchased from the Biomedical Services Unit at The University of Liverpool and culled under local and national schedule one guidelines. All procedures were carried out according to the UK Home Office regulations.

2.1.5 DNA samples

Order	Suborder	Infra order	Family	Species	Sample size	Type of sample
Primates	Anthropoidea	Catarrhini	Hominidae	*Homo sapiens*(+)	1	tissue
				Pan troglodytes (*)	48	tissue
				Gorilla sp. (°,▪)	3	tissue
				Pongo pygmaeus	2	tissue
			Cercopithecidae	*Macaca nigra* (*)	1	tissue
				Mandrillus sphinx (*)	1	tissue
				Chlorocebus aethiops (*)	1	tissue
		Platyrrhini	Cebidae	*Lagothrix spp* (**)	2	hair
				Ateles chamek (**)	1	hair
				Alouatta seniculus (**)	1	hair
				Cebus paella (**)	1	hair
				Cebus albifrons (**)	1	hair
				Saimiri sciureus (**)	1	hair
				Callicebus moloch (**)	1	hair
			Callitrichidae	*Saguinus fuscicollis*(**)	1	hair
				Callitrhix pygmea (**)	1	hair
	Prosimii		Tarsiidae	*Tarsius bancanus* (▪)	3	tissue
			Lemuridae	*Lemur catta* (*)	1	tissue
Rodentia	Myomorpha		Muridaea	*Rattus norvegicus* (+)	1	tissue
				Mus musculus(+)	1	tissue
Chiroptera	Megachiroptera Microchiroptera		Pteropodidae Phyllostomidae	*Pteropus rodricensis* (*)	1	tissue
				Carollia perspicillata (*)	1	tissue

Table 2.2 DNA samples used for this project. For this project samples from mammals included in 3 orders (Primates, Rodentia and Chiroptera) were analysed. The samples represent diverse families within each order; and were mainly obtained from animals housed in Chester Zoo, Chester, UK (*), the patronato del parquet de las leyendas (PATPAL) zoo in Lima, Peru (**), the Natural History Museum, London UK (▪) and from commercial sources (ECCAC=° and Novagen=+).

2.2 Methods

2.2.1 DNA Isolation

2.2.1.1 DNA isolation from hair strands

DNA was extracted from hair samples following the Chelex method (Walsh et al., 1991) modified by Jensen-Seaman and Kidd (2001). In brief, 5 hair strands were cut 5 mm from the root and washed in deionised water. The hair roots were placed in 1.5 ml eppendorf tubes and 250 µl of Chelex solution (section 2.1.1) was added to the tubes and incubated for 8-12 hours at 55°C. After incubation the tubes were vortex briefly and the Proteinase K reaction was stopped by immersing the samples into a water bath at 99-100°C for 8 minutes. Then, the tubes were vortexed and centrifuged at 13000 rpm for 1 minute. An aliquot of 10 µl of the supernatant was used as DNA template for each PCR reaction. The samples were stored in a freezer at -20°C.

2.2.1.2 DNA isolation from blood and tissue samples

2.2.1.2.1 TRIzol reagent method

DNA was extracted from tissue samples using the TRIzol® method of extraction (Chomczynski, 1987). In brief, 0.75 ml of TRIzol reagent (Invitrogen, UK) was added to total blood (0.125ml of blood mixed with an equal volume of dH_2O) or 100 mg of tissue. The blood was diluted because its viscosity blocked the pipette tips used. The blood samples were homogenised by pipetting up and down and the tissue was triturated and homogenised using a microtissue homogeniser (VWR Cat. No 432502). After this, the samples were incubated for 5 minutes at room temperature followed by the addition of 0.2 ml of chloroform. The tubes were shaken vigorously by hand for 15 seconds, incubated at room temperature for another 15 minutes and centrifuged at 12000x g for 15 minutes at 4°C. The centrifugation separated the

mixture into three layers. While the RNA is in the upper, aqueous phase, the DNA and proteins are in the lower, organic phenol-chloroform phase. The aqueous phase containing RNA was decanted away and an inter-phase and a phenol-clorophorm phase containing the DNA and protein were poured into a clean 1.5 ml tube. The DNA is then precipitated by the addition of ethanol (0.3 ml of 100%) and the proteins remain in the phenol-ethanol supernatant. The tubes were stored at room temperature for 2-3 minutes followed by centrifugation at 2000 x g for 5 minutes at 4°C. The supernatant was discarded and the DNA pellet formed was washed twice with 1ml of a solution containing 0.1 M sodium citrate in 10% ethanol. After each wash, the DNA pellet was stored for 30 minutes at room temperature followed by centrifugation at 2000xg for 5 mins at 4°C. Finally, the DNA pellet was washed in 75% ethanol, incubated for 10-20 minutes at room temperature and centrifuged at 2000g for 5 minutes at 4°C. The DNA pellet was air dried for 5 minutes and redissolved in 8 mM NaOH by pippeting up and down to achieve a final concentration of 0.2-0.3 µg/µl of genomic DNA.

2.2.1.2.2 Ethanol precipitation method

DNA preparations using the TRIzol reagent often carried over impurities (cellular debris). Thus, the DNA was isolated 2nd time using the ethanol precipitation method, which allowed the removal of unwanted salts and contaminants. The protocol is as follows: 200 µl of the DNA solution were transferred into a 1.5 ml eppendorf tube and 20 µl of sodium acetate buffer (section 2.1.1) and 400 µl of ice-cold 100% ethanol were added. The tubes were placed in a freezer (at -20°C) for at least 1 hour. The samples were then centrifuged at 13000 rpm for 15 minutes in a pre-cooled bench top centrifuge (4 °C). The supernatant was carefully removed with a 1 ml pipette. Ice-cold 70% ethanol (200 µl) was added to the samples and then centrifuged for 5

minutes at 13000 rpm at 4°C. The supernatant was removed using a 200 µl pipette, and allowed to air dry. Finally, the pellets of DNA were resuspended in 100-200 µl of DNAse-free water. The concentration of DNA was calculated using a spectrophotometer as described in section 2.2.5.

2.2.2 Total DNA amplification using Genomi Phi

The DNA isolated from hair samples was often of very low concentration and unsuitable for direct PCR. Such DNA samples were amplified using the GenomiPhi amplification kit (Amershan Biosciences). The GenomiPhi kit utilizes bacteriophage Phi29 DNA polymerase to amplify single- or double-stranded linear DNA templates via a "strand displacement" reaction. The genomic DNA is mixed with random hexamer primers, and after these have primed the other reaction components (Phi29 DNA polymerase, deoxynucleotide triphosphates, and buffers) are added. This reaction mixture is incubated, and during this time the reaction converts the nucleotides into a high molecular weight fragment copies of the template DNA. The accuracy of this DNA replication is high because the polymerase used has high proofreading activity.

The protocol of the amplification is described as follows: first 1 µl of the DNA solution was mixed with 9 µl of the sample buffer in a microcentrifuge tube (200 µl) and the mix was incubated at 95°C for 3 mins to denature the DNA. Subsequently, the tube was cool to 4°C on wet ice. In a separate tube, 1 µl of the enzyme mix and 9 µl of reaction buffer were mixed, and this was added to the previously ice cooled sample. The contents of both tubes were mixed on ice and then immediately placed in a PCR machine at 30°C for 16-18 hours to allow the amplification reaction to proceed. Finally, to stop the action of the Phi29 polymerase, the samples were incubated at 65°C for 10 minutes and cooled to 4°C. The samples were ethanol precipitated and the

concentration of the newly amplified DNA sample was calculated by visualisation using gel electrophoresis and spectrophotometry (detailed in section 2.2.5).

2.2.3 PCR

2.2.3.1 PCR Primer design

If the flanking sequences of a VNTR region were unknown, PCR primers were designed to prime on evolutionary conserved regions (non-coding or exonic) flanking the VNTR. The conserved areas were determined using alignments generated by the conservation tool of the Human genome BLAT search from the UCSC browser (http://genome.ucsc.edu/cgi-bin/hgBlat) which aligned mammalian and vertebrate genomes. To avoid primer mispairing, primer sequences were chosen from regions that presented ≥ 85% conservation across mammalian sequences (see Figure 2.1). PCR amplification using such "conserved" primers produced long PCR products (ranging from 1.5-2.5 kb long). Once this first step of PCR amplification was performed, primers internal to the long set, design to produce smaller products were designed. A concentration curve was performed for each primer pair to determine the optimum primer concentration during PCR. For nested PCR (section 2.2.3.2), at least one internal or nested primer was used. Routinely, restriction enzyme sites of *Xho*I and *Acc65*I were inserted at the 5' end of the primers (newly designed and previously published) for further genetic manipulation. Nested primers were designed to prime human genomic DNA. Sequences for some nested primers were obtained from previous publications.

As the PCR technique could be subject to misteakes in the amplification of VNTR and ECR fragments, a polymerase with high fidelity was employed. Also, the

generated sequences were compared to homologous sequences found in the literature (whenever available) to verify the integrity of the newly amplified.

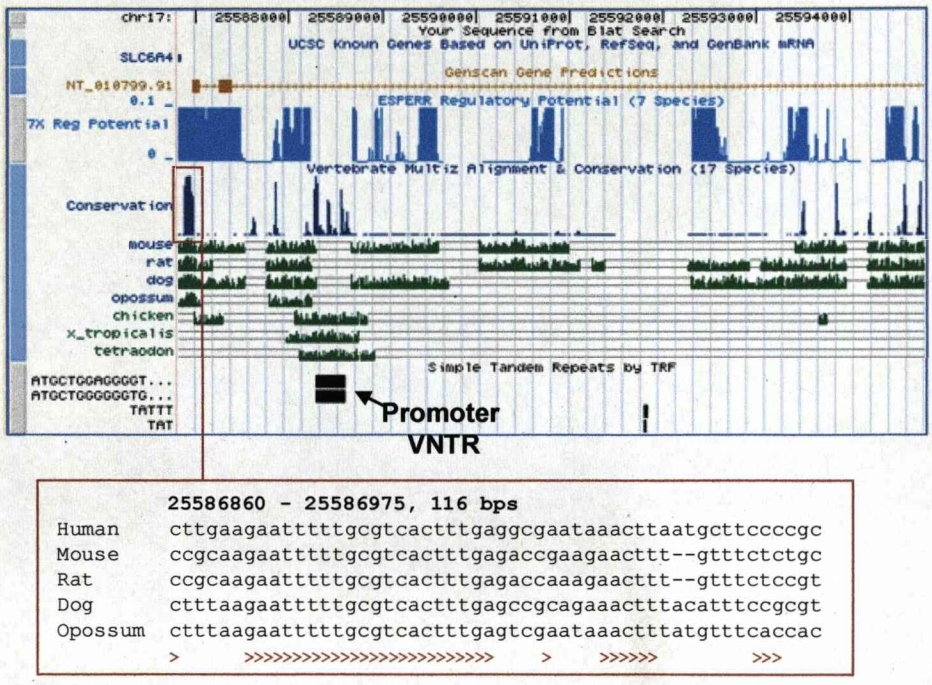


Figure 2.1 Example of design of primers for amplification of a VNTR region. Multiple sequence alignment using the conservation tool of the BLAT search engine (from the UCSC genome browser) reveal evolutionary conserved area across mammals (*H. sapiens*, *M. musculus*, *R. norvegicus*, *Canis familiaris* and *Monodelphis domestica* genomes) surrounding the promoter VNTR of the SLC6A4 gene (peak of conservation is marked by a red rectangle in graph). The sequence fragment exhibiting high conservation is encircled in red. Most conserved sections in the multiple alignment were used to design primers for PCR of VNTR in all mammals and primate species analysed. Candidate primers were located in areas of greatest conservation (undelined by >).

2.2.3.2 Nested PCR

A second amplification round (nested PCR) was performed using primers, which primed closer to the VNTR region. An aliquot (1 µl) from the reaction of the first round of PCR was used as a template for the nested PCR amplification. Whenever available, previously published primers used to amplify VNTR in non-human primate DNA were used for the nested PCR. In general, repetitive sequences with partial homology to the repeats found within the VNTR expanded beyond the VNTR locus, thus the nested primer sequences were located 150-200 bp upstream and downstream of the VNTR of interest.

2.2.3.3 PCR mix

The VNTR regions were amplified by PCR. The primers and the PCR conditions used in specific reactions are detailed in appendix 1. Commonly, the 50µl reaction comprised the DNA sample (10-100 µg), 2 mM $MgCl_2$, 1X NH4 PCR buffer, 1 unit of Diamond DNA polymerase (Bioline), 1-4 µM of each primer and 200 µM of each dNTP. Due to the high GC content which contributes to formation of secondary DNA structure within the VNTR fragments, detergents such as DMSO (5% v/v) or Betaine (0.1 M) were added. Also, dGTP was replaced by 7-deaza dGTP:dGTP (50:50, v:v) for PCR amplifications of the VNTR in the third exon of the DRD4 gene. The DNA template was denatured separately for 1 minute at 99 °C prior to addition of the other reagents. Primers and specific conditions for each PCR performed are detailed in appendix 1. The PCR reactions were performed in a Personal thermal cycler PX2 (Thermo Scientific).

2.2.4 Gel electrophoresis.

For visualization of the PCR amplicons, 15 µl of PCR reaction was mixed with 3 µl of 6x loading buffer (section 2.1.1) and then loaded in agarose gel (1- 1.5% agarose depending on the size of the PCR product, 1x TBE stained with Ethidium Bromide final concentration 0.5µg/ml). In average, most samples separated sufficiently when the samples were run in an electrophoresis tank for 1 hour at 0.5V/cm, after this period bands migrated enough to be compared with a DNA marker (100 bp or 1 kb ladders Promega, Madisson catalogue numbers G8291 and G7541 respectively). Then, the electrophoresis gel was removed from the tank and placed into a long wave UV transIluminator (Alpha Innotech, San Leandro, CA) for visualization of the PCR bands. Gel electrophoresis was also used for determining the concentration of plasmid DNA obtained from minipreps or maxipreps (section 2.2.10). Before loading DNA in the agarose gel, plasmid DNA was linearised by enzymatic digest and diluted down to compare with an appropriate DNA mass marker (Mass ruler, Fermentas; Cat No SM0383). DNA concentration was calculated based on intensity of the band and confirmed by spectrophotometry (section 2.2.5).

2.2.4.1 Gel purification

After electrophoresis, bands corresponding to products of the predicted size were cut out of the agarose gel using clean blades and the DNA was extracted from the gel using the QIAquick Gel extraction kit (QIAGEN Cat. No.28706). The gel slices were weighed and three volumes of the buffer QG were added to the tubes followed by incubation at 50°C until the gels were completely dissolved. One volume of isopropanol was added to clean PCR products smaller than 500 bp. The sample mix was poured into the QIAquick spin column, placed in a 2 ml tube and centrifuged at 13000 rpm for 1 minute. The supernatant was discarded and the column was washed

with 0.75 ml of PE buffer by centrifugation for 1 minute. The flow through was discarded and the column was centrifuged again for 1 minute at 13000 rpm. The column was placed into a clean eppendorf tube and 30 µl of dH_2O was added to the centre of the column, incubated for 1 minute and finally centrifuged for 2 minutes at 13000 rpm, where the eluted volume contained the plasmid DNA. Due to loss of DNA during the gel purification process an aliquot of the purified DNA was run in an electrophoresis gel to determine the final concentration.

2.2.5 Measurement of DNA concentration and quality by spectrophotometry

Exact DNA concentration was calculated by spectrophotometry. The UV spectrophotometer (Jenway Genova Life Science Analyser catalogue number 636 031) was calibrated using 100 µl dH_2O (solvant) as blank. After calibration, 2 µl of DNA preparation was diluted in 98 µl of dH_2O, placed into a quartz cuvette (200 µl) and placed in the cell holder for the determination of concentration, using the following formula:

Original concentration = O.D value "X" (at wavelength WL of 260 nm) x 50 ng/ml x dilution factor,

Where 1 O.D. at 260 nm for double-stranded (ds) DNA equals 50 ng/ml of dsDNA. The spectrophotometer can measure at two wavelengths (WL1 = 260 nm, and WL2 = 280 nm). The WL1 (260 nm) measures the DNA concentration as well as the RNAs if exist. The WL2 (280 nm) measures the protein contamination. The ratio of WL1 to WL2 should exceed 1.80 for best results or good purification of DNA preparation.

2.2.6 Cloning of VNTRs

2.2.6.1 Addition of 3' overhangs to a blunt end PCR fragment

The enzyme used in PCR (Diamond DNA polymerase, Bioline Cat. No. BIO-21059) generated blunt ended fragments, which required modification before ligation to intermediate plasmid vectors (pGEM-t), designed to make use of overhangs generated by *Taq* polymerase. Thus, the gel-purified fragments containing the VNTR of interest were modified using an A-tailing procedure which created an overhang of adenine nucleotides at the 3' end of the fragment, complementary to the thymidine overhang found in the pGEMt. In brief, the procedure included the following:

1- 7 µl of the purified PCR product

dATP final concentration of 0.2 mM

5 units of *Taq* DNA polymerase

Taq $MgCl_2$ free buffer (final concentration 1X)

25 mM $MgCl_2$

Destilled H_2O was added to the mix to achieve a final volume of 10 µl followed by incubation at 72 °C for 15 to 30 minutes. After this incubation period, the tubes containing the reaction were placed in ice to halt the reaction.

2.2.6.2 Ligation of compatible termini created by restriction enzymes

After addition of A-overhangs, the fragments were ligated into an intermediate vector pGEM-t (Promega). Routinely, the ligation reaction comprised:

2 µl of A-overhang reaction (section 2.2.6.1)

50 µg of the linearised plasmid (pGEM-t)

200 units of T4 DNA ligase (NEB M0202S)

T4 DNA ligase buffer 1 X final concentration and dH_20 to a final volume of 10 µl.

In most cases, the ratio of the concentration of the VNTR fragments: plasmid vector used for the ligation was 3:1 or 5:1. The mix was first incubated at 14-16 °C for 3 hours and then at 4 °C overnight.

To clone the VNTR fragments into a luciferase expression vector (PGL3p), the fragments were excised from the intermediate vectors (pGEM-t) by enzymatic digest using the restriction enzyme sites introduced by the nested primers (*XhoI* and *Acc65I*). This facilitated directional cloning and permitted their ligation to linearised expression vector pGL3p (by enzymatic digest with *XhoI* and *Acc65I*) following the ligation procedure described above.

2.2.7 Transformation of chemically competent *E. coli* cells.

Once the generation of recombinant plasmid DNA was confirmed by enzymatic digest (see Figure 2.3), it was required to generate large amounts of plasmid DNA for cloning and later molecular manipulation. Strains of competent *E. coli* cells (DH5-α, Invitrogen-Gibco BRL cat No 18265-017, TOP 10 Invitrogen-Gibco BRL cat No C4040-10) were used for transformation with intermediate and final plasmid vectors respectively. Briefly, 50 µl of competent cells were thawed on ice, 1-10 ng of plasmid DNA (1-4 µl of the ligation reaction) were added and subsequently incubated on ice for 15 min. The cells were heat shocked in a water bath for 30 sec at 42°C and placed on ice for 2 minutes. 500 µl of room temperature LB broth (for DH5-α cells) or 450 µl of SOC solution (for TOP-10 cells) was added to the tubes containing the transformation and were incubated at 37 °C for 1 hour in a shaker incubator (225 rpm). 50-200 µl of the transformation mix was spread on LB

agar plates containing 50-100μg/ml ampicillin (for pGEM-t and pGL3p), flipped over and incubated at 37°C in an incubator for 16 to 18 hours.

2.2.8 Blue/White screening of transformants

To help identify positive transformants of the ligation between the PCR products containing the VNTR regions and the intermediate vector (pGEM-t) Xgal (5-bromo-4-chloro-3-indolyl-bD-galactoside dissolved in dimethylformamide, Promega No. V3491 50mg/ml) and IPTG (Isopropyl β-D-1-thiogalactopyranoside, BIO-37036) were plated and spread onto the selective agar plates. After spreading these solutions, the agar plates were placed in an incubator for 30 minutes at 37°C prior use. Successful ligation resulted in the integration of the VNTR fragment into the plasmid's multiple cloning site, which was located within the amino terminal fragment of the β-galactosidase gene. After induction caused by IPTG, bacteria not carrying the plasmids formed blue colonies in the presence of the chromogenic substrate X-gal; on the other hand, bacteria carrying the recombinant plasmid remained white. Blue coloration of the colonies developed after 17 hours of incubation at 37 °C. For each transformation, 10-30 single white colonies were picked and used to grow 5 ml LB cultures for DNA minipreps. I then confirmed through diagnostic enzymatic digest the insertion of the VNTR/ECR fragment.

2.2.9 *E.coli* colony PCR

Transformation of TOP10 cells with pGL3p often produced abundant colonies (≥100 per 50 μl of transformation reaction). Thus, colony PCR was used to rapid screen many colonies simultaneously (Figure 2.2). In brief, single colonies were picked freshly streaked L-B agar selective plates with a yellow pipette tip (200 μl) and swirled into a microcentrifuge tube (200 μl) containing 25 μl of distilled water.

Tubes were vortexed until water became cloudy. Then, tubes were immersed in a water bath (heated to 100 °C) for 2 minutes and centrifuged at 13 000 rpm for 2 minutes in a bench top microcentrifuge. Finally, 20 µl of the supernatant was transferred into a fresh tube and 5-10 µl were used as template per 50 µl of PCR reaction.

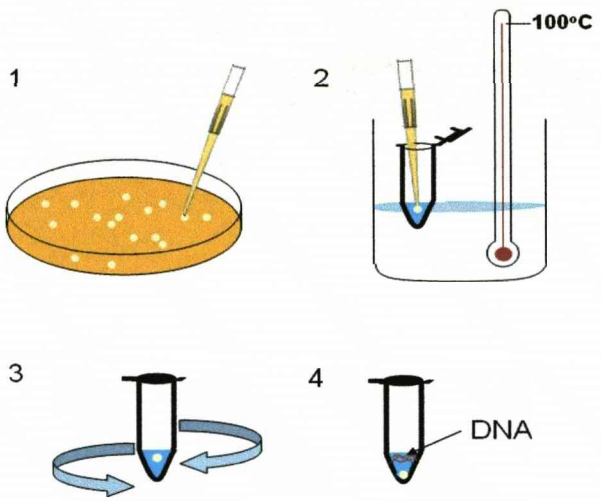


Figure 2.2 *E. coli* colony PCR. DNA template was obtained immediately from *E. coli* colonies for quick confirmation of insertion of VNTR/ECR fragment into plasmid vector. Colonies were picked with a pipette tip (1), swirled into an eppendorf tube with dH_2O (2), placed in a water bath (100°C) for 2 minutes (2) and centrifuged afterwards (3). The supernatant containing the plasmid DNA was used for PCR (4).

2.2.10 Small scale preparation of plasmid DNA

Routinely, after transformation of competent cells with ligation reactions, small preparation of plasmid DNA were required to verify insertion, position of insert into the plasmid vector and subsequent cloning applications. In brief, single colonies were picked from a freshly streaked selective plate (ampicillin or kanamycin), and were used to inoculate a starting LB culture (5 ml) containing ampicillin (100 µg/ml) or Kanamycin (25 µg/ml) and left to grow for 12 hours. After the growth period, a 500 µl aliquot was mixed by repeated inversion with an equal amount of glycerol (at 50% with dH_2O, v/v) and stored at -80 °C for further manipulation. The remaining left

4.5 ml of culture were routinely used for mini preparation of plasmid DNA using the QIAprep spin Mini prep kit (Qiagen cat. 27106). The cultures were subsequently centrifuged at 13000 rpm in a bench top centrifuge and the supernatant discarded. The bacterial pellet formed was resuspended in 250µl of the resuspension buffer (P1) until no clumps of cells were observed. 250µl of lysis buffer (P2) was added to the tubes, inverted 6 times and incubated at room temperature for 5 minutes. An aliquot (350 µl) of the neutralization buffer N3 was added to the tubes and mixed well by inversion. The lysates were centrifuged at 13000 rpm for 10 minutes and the flowthrough was poured into a QIAprep spin column. The columns were placed into 2 ml microcentrifuge tubes and centrifuged for 1 minute at 13000 rpm. After this, the columns were washed once with 750 µl of PE buffer and centrifuged at 13000 rpm for 1 minute, after centrifugation, the container tube was replaced with a clean eppendorf tube. Finally, to elute the plasmid DNA, 30 µl of dH_2O were added to the centre of the column, left to rest for 2 minutes and centrifuged for 2 minutes at 13000 rpm. The concentration and integrity of the plasmid DNA were analysed by using gel electrophoresis and spectrophotometry. The direction and insertion of the VNTR fragment was confirmed through diagnostic enzymatic digest (section 2.2.12) colony PCR (section 2.2.9) and sequencing.

2.2.11 Large scale preparation of plasmid DNA

Transfection experiments required large quantities of recombinant plasmid DNA, thus large preparations of DNA were produced using a Plasmid Maxi Kit (QIAgen, Cat. No 12163). In brief, a starting culture (5ml of LB) which contained selective antibiotic (ampicillin 100 µg/ml or Kanamycin 25 µg/ml) was inoculated with bacteria carrying the plasmid of interest and this was incubated for 8 hours at 37 ^{o}C in a shaker incubator (225 rpm). After this period, 0.5 ml of the starter culture was

added to 100 ml of LB containing ampicillin (100 μg/ml). The cultures were incubated for 12-16 hours at 37°C in an incubator shaker (225 rpm). After incubation, the cultures were centrifuged in a table-top centrifuge at 3000 rpm for 15 minutes at 4 °C. The supernatant was discarded and the formed bacterial pellets resuspended in 10 ml of the re-suspension buffer P1. Then, pellets were lysed by adding 10 ml of lysis buffer P2 mixing fully by inversion (4-6 times) followed by incubation for 5 minutes at room temperature. 10 ml of the pre-chilled precipitation buffer P3 was added to the tubes containing the lysates, inverted 6 times and incubated 20 minutes on ice. The tubes were then centrifuged at 20000x g for 30 minutes at 4 °C and the supernatant containing the plasmid DNA was transferred to a clean 50ml Oakridge tube. To completely remove cell debris, the supernatant was centrifuged again at 20 000x g for 15 minutes at 4 °C. The supernatant containing plasmid DNA was applied to an equilibrated column (QIAGEN tip 500) and allowed to enter by gravity flow. The column was washed twice with 30 ml of buffer QC. After the washes, elution buffer QF (15 ml) was added and the eluate collected in a clean Oakridge tube. DNA was precipitated by adding 10.5 ml of isopropanol (v/v) at room temperature to the eluate, mixed by inversion and centrifuged at 15,000xg for 30mins at 4°C. Isopropanol was carefully decanted without disturbing the DNA pellet formed. The DNA pellet was washed with 5 ml of 70% ethanol and centrifuged at 15,000g for 10mins at 4°C. Supernatant was carefully decanted and tube was inverted on tissue paper for 5-10 mins to air dry the DNA pellet. DNA was redissolved in 200 μl of dH_20. The integrity and concentration of the plasmid preparation were analysed by agarose gel electrophoresis and spectrophotometry (section 2.2.5).

2.2.12 Analytical restriction enzyme digests

Restriction enzymes were used to verify the insertion and position of the VNTR fragments into the plasmid vectors (Figure 2.3). Routinely, fragments cloned into pGEM-t vector multiple cloning site were analysed by the use of *Nco I* and *SacI*. In addition, enzymes cutting a maximum of twice inside the VNTR fragment were used to determine the direction the VNTR fragment fragment had been cloned. To verify the position of the insert in the pGL3p vector, recombinant plasmid DNA was digested with *Nco I* and *Not I* (Promega, UK). The fragments were visualized after gel electrophoresis in a UV light transIluminator.

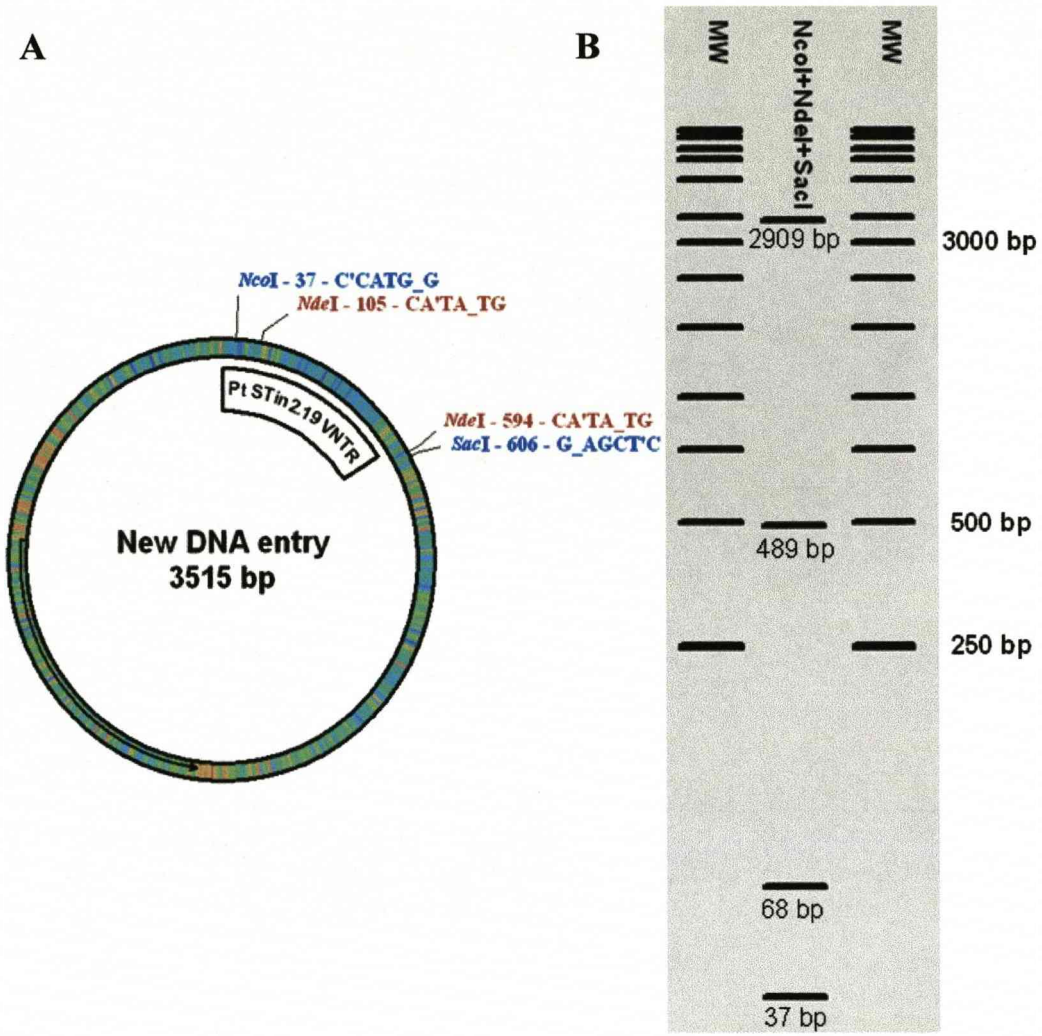


Figure 2.3. Diagnostic digest to test insertion and direction of VNTR region into plasmid vector. Example of diagnostic digest of *P. troglodytes* STin 2.19 VNTR fragment cloned into pGEM-t. Positive clones (VNTR fragment inserted into the plasmid multiple cloning site are represented by plasmid map (A) produced 4 bands when cut using *Nco I, Sac I* (restriction sites in the multiple cloning site) and *Nde I* (cut twice within the VNTR fragment B). Electronic representation of enzyme digest was generated using pDRAW. MW= molecular weight

2.2.13 Isolation of total RNA

To correlate the transfection data of the VNTR supported reporter gene with the endogenous gene expression, it was necessary to know whether which neurotransmitter genes were endogenously expressed in the cell culture models used. The presence of the mRNA transcripts was confirmed by Reverse transcriptase PCR

(RT-PCR). For this application, it was necessary to produced complementary DNA (cDNA) from the total RNA isolated from cell cultures and tissues. Total RNA was isolated from frontal cortex and midbrain sections of male Wistar albino rats immediately after culling. In brief, the dissected tissue (≈75 mg) was snap-frozen by immersion on liquid nitrogen followed by homogenization (using a homogenizer VWR Cat. No 432502) in 0.75 ml (ratio 1 tissue (w):10 TRIzol (v)) of TRIzol reagent (Invitrogen) and then incubated for 5 minutes at room temperature. Chloroform (0.2 ml) was added to the homogenized samples, mixed by vigorous handshaking for 15 sec and centrifuged at 12000 x g for 15 mins at 4 °C. After centrifugation, the mixture separated into 3 phases, the RNA remained in the upper aqueous phase. To precipitate the RNA, the aqueous phase was transferred into a clean eppendorf tube where it was mixed with 0.5 ml of isopropanol (Sigma-Aldrich W292907). The samples were incubated for 10 minutes at room temperature and after incubation centrifuged at 12000 xg at 4 °C. After centrifugation, a pellet was formed at the bottom of the tube. The supernatant was removed carefully and the pellet was washed with 0.75 ml of 75% ethanol, vortexed and centrifuged at 7500 xg for 5 mins at 4 °C. Finally, the supernatant was removed and the RNA pellet was air dried for 5 minutes and reconstituted in RNAse-free water. The RNA solution was run in an agarose gel where it was observed as two bands (28S and 18S) indicative of intact ribosomal RNA. The concentration of the RNA solution was assessed by spectrophotometry.

2.2.14 Synthesis of cDNA by reverse transcription polymerase chain reaction (RT-PCR).

From the RNA isolated in the previous section, cDNA was produced by reverse transcription. Briefly, 0.6 μg of total RNA was used in the reaction. The RNA was bound to 1 μl of oligo dT in a 5 minutes reaction at 70°C, followed by 4°C for 5

minutes. This step generated single stranded DNA from the RNA. The complementary strand of DNA was produced by the reverse polymerase reaction as follows: a reaction mixture of 4 µl of reverse transcriptase buffer, 2 µl of dNTPs, 0.5 µl ribonuclease inhibitor, 2.4 µl 25mM magnesium chloride, 5 µl dH_2O and 1 µl reverse transcriptase was added to the RNA mix. The reverse transcription reaction was carried out at 25° C for 3mins, 42° C for 60 mins and 70° C for 15mins. Negative controls (reverse transcriptase missing) were also included in the experiment. The PCR primers used are listed in appendix 2.

2.2.15 Cell culture

2.2.15.1 Dissection of frontal cortex and midbrain from neonate Wistar rats

Rats aged 2-7 days old were culled in a gas chamber (CO_2). The head was severed from the body using a sharp razor blade applied in the dorsal aspect of the neck area (Figure 2.4a). Using the blade, a midline incision was made following the longest axis of the head; care was taken not to damage the underlying brain tissue by using excessive force since the bone plates of the neonate rat skull have not finished ossification. The flaps of skin were reflected laterally and any connective tissues (fascia, fat) were removed. The skull was held in place by curved forceps (Fisher Scientific catalogue number DKC-790-D) and using micro scissors (WPI cat No 501778) another longitudinal incision starting at the bregma point was made along the sagittal suture of the skull, keeping the tips of the scissors pointing upwards to avoid damaging brain tissue. The skull plates were held by curved forceps and the dura matter, which was still attached to the skull and brain was removed using microscissors. The scoop-end of a spatula (Fisher scientific, Cat No 3006) was placed between the ventral surface of the brain and the inside of the base of the skull. The

spatula was then carefully moved from side to side to cut the underlying optic nerve tracts, releasing the brain. The brain was removed using the spatula and placed in a petri dish filled with pre-chilled dissection solution (section 2.1.3.2.3). Once in the petri dish, the frontal cortex was dissected out using a clean scalpel by inserting the blade with a 50° inclination pointing towards the root of the optical bulb (see Figure 2.4a). The temporal cortex was obtained by making a round shallow incision (5mm diameter) with the scalpel (Figure 2.4a). To obtain midbrain tissue sections, the brain was then flipped over, so that the ventral surface faced upwards. Then, using a blade positioned at 90 ° (3 mm deep) an incision was made, 3 mm anterior to the borderline drawn between the cerebral hemispheres and the brainstem. Subsequent incisions were made to obtain a square section 3mm per side, which centre was defined by the borderline drawn between the brainstem and the cerebrum and the longitudinal axis of the brain (Figure 2.4b). Finally, the dissected frontal cortex and midbrain tissues were placed into a 30 ml flask containing dissection solution for preparation of disassociated cultures (section 2.2.15.2).

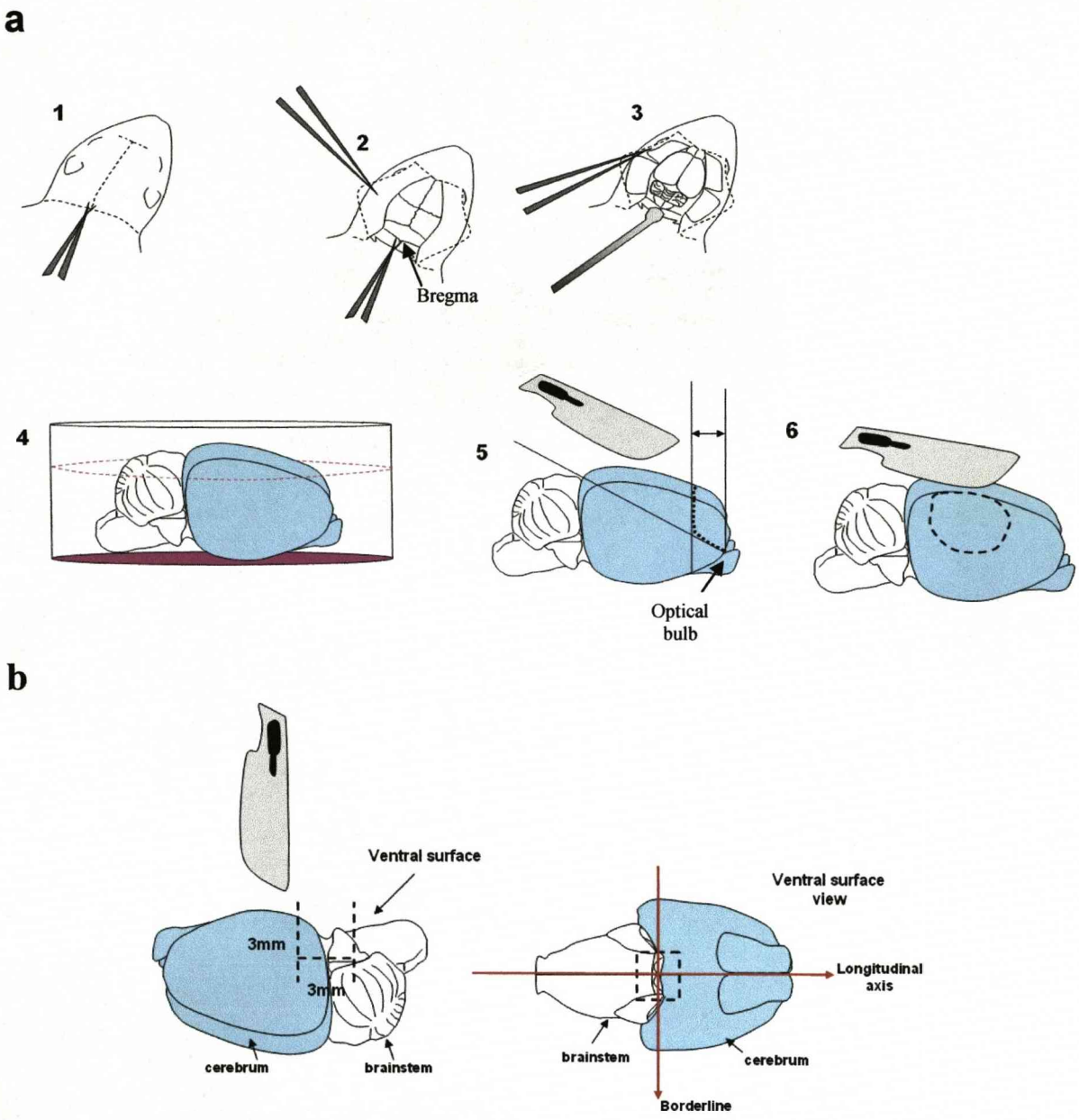


Figure 2.4 Dissection of neonate rat cortex and midbrain tissue (a) Dissection of temporal and frontal cortex from the neonatal rat brain. After culling, skin was cut by longitudinal incision (1) and reflected laterally, the skull plates were dissected away similarly (2). Once exposed, brain was removed using a bent ended spatula (3) and placed in chilled dissection solution (4). Dissection of frontal cortex (delineated by dotted line) was performed by directing the scalpel in a 50° angle towards the root of the optic bulb (5). The temporal cortex was obtained by making a shallow incision (2mm deep) on both cerebral hemispheres (6). **(b)** Dissection of the rat midbrain. Once removed out of the skull, the brain was flipped over and a square section (3mm each side) was dissected out. Cortex coloured in light blue and the brainstem is white.

2.2.15.2 Preparation of cortical and midbrain cultures

Cortex and midbrain tissues were stored in a flask containing dissection solution and centrifuged at 1000 rpm for 5 minutes at room temperature in a bench top centrifuge. The supernatant was replaced with 3 ml of trypsin/EDTA solution (2.5 g porcine trypsin and 0.2 g EDTA·4Na per litre of HBSS, Sigma-Aldrich Ltd cat No T4049) and placed in an incubator at 37°C for 20mins. The tissue was then centrifuged at 500 rpm for 5 minutes at room temperature; after which the trypsin solution was decanted and replaced by fresh pre-warmed (37°C) culture medium I containing Penicillin/streptomycin (section 2.1.3.2.1). The tissue was then centrifuged at 500 rpm for 3 minutes at room temperature, this procedure was repeated three times. The resulting pellet was dissociated in 5 ml of culture medium I (penicillin/streptomycin) using 2 Pasteur pipettes with pores of decreasing diameter until the cell suspension was homogeneous and the solution appeared turbid. The resulting suspension was passed through a 0.70μm falcon cell strainer (VWR) to remove debris, centrifuged at 1000 rpm for 5 minutes and cells are resuspended in 5 ml of culture medium I (without antibiotics). The disassociated cells were counted under a contrast microscope using a cell counting chamber (10^5 per well). The cells were then plated into poly-D-lysine coated (section 2.2.15.5) 24 well plates for transfection 24 hours later. Poly-D-Lysine was used as this substance will create a matrix for the better adherence of neuronal cultures to the floor of the culture flask. After 7 hrs, medium I was removed and replaced with 1 ml/well of medium II. Prior to transfections, the medium II was renewed.

2.2.15.3 Culture of JAr cells

JAr cells were cultured at 37 °C, 5% CO2, in medium described in section 2.1.3.1.1. The cells were cultured in monolayer of cells adherent in T75 ml flasks and

the culture media was changed every other day. Cells were split when 70–80% confluence was achieved.

2.2.15.4 Culture of SN4741 cells

SN4741 is a mouse substantia nigra-derived dopaminergic neuronal cell line and were cultured at 33°C, 5% CO_2, in culture medium described in section 2.1.3.1.2. The maintanece of the cells was similar as described for JAr cells.

2.2.15.5 Coating tissue culture plastics

Tissue culture plates used to grow primary brain cultures were coated with poly-D-lysine (100 µg/ml) which was prepared as a stock solution at concentration of 10mg/ml, aliquoted and stored at -20°C under sterile conditions. Prior to transfection, Poly-D-lysine was dissolved in sterile dH_2O and 200 µl (12.5 $\mu g/cm^2$) were added to coat each well. The 24 well-plates were placed in a 37°C incubator for 1-17 hrs to allow the poly-D-lysine to adhere to the surface of the wells. The plates were washed twice with PBS (1X) prior to plating of cells to remove any traces of poly-D-lysine

2.2.16 Delivery of luciferase constructs into cell lines and rat cortical cultures

Plasmids carrying the VNTR/ECR fragments were delivered into the cells using either ExGen 500 (Fermentas), TRANSFAST reagent (Promega) or the nucleofector device (Nucleofector I, AMAXA Biosystems). For each transfection performed, a vector pmLuc-2 containing a minimal TK promoter followed by an optimized *Renilla* luciferase (Rluc) cDNA (Novagen) was co-transfected to normalise transfection efficiency (ratio 100: 1, VNTR construct: pmLuc-2 plasmid).

2.2.16.1 Exgen 500

This reagent was used for transfection of plasmid DNA into primary cultures of neonatal rat cortex. ExGen 500 (Fermentas) is a polymer with high cationic-charge density, which forms complexes with DNA that are internalised by endosomes. Rupture of the endosomes into the cytoplasm allows the translocation of the DNA into the nucleus (Fermentas). Briefly, 1 μg of VNTR/ECR construct and 0.01 μg of the internal control plasmid pmLuc-2, were diluted in 100 μl of 150 mM NaCl, followed by gentle vortexing and brief centrifugation. After this, 3.3 μl of ExGen 500 were added per 1 μg of DNA used. The solution was immediately vortexed for 10 seconds and incubated for 10 minutes at room temperature. 100 μl of the ExGen 500/DNA mixture was added to each well (the volume of the mix ExGen 500/ DNA mixture represented 10% of the total volume of the culture medium) and gently rocked back and forth and from side to side to achieve even distribution of the complexes. The plates were then centrifuged for 5 min at 280xg at room temperature and finally incubated at 37°C for 48 hours in a humidified 5% CO_2 incubator. After this period, the cells were harvested for luciferase assay.

2.2.16.2 Nucleofector device

Delivery of plasmid DNA into primary cultures was also achieved using the Nucleofector I device (Amaxa GhmB, Koeln, Germany). The transfection protocol was optimised for rat hippocampal/cortical neurones. Midbrain tissue was dissected and disassociated as described in sections 2.2.15.1 and 2.2.1.5.2. The neuronal enriched cultures were resuspended in the Neuron Nucleofector Solution (supplied by the manufacturer) prepared in 100 μl at room temperature to a final concentration of 4 - 5 x 10^6 cells/100 μl. 3 μg DNA (0.5 μg of TF per 1 μg of luciferase plasmid for co-transfections) were mixed with 100 μl of cell suspension. The mixed sample was transferred into a cuvette (Amaxa) making sure the sample covered the bottom of the

cuvette and avoiding the formation of air bubbles. The cuvette was closed with the blue cap, inserted into the cuvette holder of the nucleofector device and rotated to final position. Program O-03 was selected and executed. The samples were removed from the cuvette immediately after nucleofection had taken place and placed into eppendorf tubes containing 500 µl of pre-warmed (to 37°C) medium I using plastic pipettes provided in the kit. The samples were transferred (1ug of luciferase plasmid per well) into the prepared 24-well plates previously coated with poly-D-lysine. Cells were incubated in a humidified 37°C/5% CO_2 incubator. 24 hours later the medium I was replaced with 1 ml of culture medium II. After 48 hours, cells were harvested for luciferase assays.

2.2.16.3 TRANSFAST

The TRANSFAST reagent was used for transfection of plasmid DNA into cell lines JAr and SN4741. TRANSFAST (Promega, Madison) is a complex formed by a synthetic cationic lipid and a neutral lipid (DOPE). The lipid complex associates with the DNA and similarly as for ExGen, it is introduced into the cell by endocytosis and later released into the cytoplasm allowing its passage to the nucleus (Promega Madison, technical bulletin TB260). The day before transfection, the TRANFAST reagent was re-suspended in dH_2O and placed in a -20 °C to freeze; simultaneously, cells were plated to achieve 60% 70% confluency at transfection. Immediately before transfection, the culture medium was replaced by serum free medium. Plasmid DNA (1 µg/well of luciferase construct and 0.01 µg/well of pmLuc-2) was diluted in 200µl of serum free culture medium in an eppendorf tube and mixed with TRANFAST (1 mM final concentration, 21 µl/1µg of DNA, ratio 2:1 TRANFAST: plasmid DNA), immediately followed by brief vortexing. The mixture was incubated for 10 to 15 minutes at room temperature. The culture medium was removed from the 24 well

plate and the transfection mixture (200 µl) was added to the wells and returned them to the humidified 37°C, 5% CO_2 incubator for 1 hour, after which 800 µl of culture medium containing serum was added to each well. The plates were returned to the incubator for 48 hours and after this period the cells were harvested for luciferase assay.

2.2.17 Co-transfection/Co-nucleofection experiments

To asses the potential regulation of VNTRs/ECRs, the full length (human) expression constructs for Sp1, CTCF and YB-1 proteins were transfected into cell lines and primary cultures of cortex simultaneously with the VNTRs/ECRs constructs (1/0.5 µg of Sp1 and 1 µg of YB-1 and CTCF and per 1 µg of VNTR/ECR construct). The constructs were co-transfected using TRANSFAST or ExGen 500 or nucleofected (using the nucleofector device) following protocols described above. In co-transfection experiments the amount of plasmid DNA transfected into the cells was maintained constant. For this, per every 1µg of expression vectors co-transfected with the VNTR or ECR constructs, equal amounts of inocous DNA (pGL3b) were transfected with VNTR constructs when basal levels of transcription were assesed.

2.2.18 Analysis of gene expression *in vitro*

2.2.18.1 Luciferase assay

Analysis of the amount of luciferase protein activity produced by the transfected plasmids was estimated using the Dual Luciferase Assay kit (Promega, Madison Cat. No E1500) on extracts of transfected cells. Briefly, cell extracts were obtained as follows: culture medium was removed and wells were washed twice with PBS. Passive lysis buffer (40 µl of 1x/well [24 well plate]) was added to the wells and incubated for 15 minutes on a rocking platform. At the end of this period,

supernatants containing lysed cell extracts were aspirated with a pipette. The supernatants (20 µl) were plated on a 96 well plate and transferred into the Glomax 96 microplate luminometer (Promega). Firefly luciferase reagent (100 µl containing the luciferase assay substrate) and sea pansy luciferase reagent (100 µl, containing the sea pansy luciferase substrate) were automatically injected into each well to calculate luminescence intensity. The sea-pansy luciferase substrate solution was added to each sample to determine the protein production of the internal control (pmLuc-2) to normalise for transfection efficiency in case the number of cells or the efficiency of the transfection varied from well to well. The calculated luminescence was processed by the Stingray 2.0 software and a readout was produced.

2.2.19 Phylogenetic analysis

To infer the evolutionary history of the different VNTR/ECR studied, I conducted phylogenetic analysis or cladistics, following the principle of parsimony. The phylogenetic analysis assesed the relatedness of ECR or VNTR sequences based on the putative TFBS found whithin the sequences. These TFBS were considered as valid characters for the study of the evolution of these *cis* regulatory domains as binding sites for TFs are the evolving units within these domains. The TFBS data obtained from the sequences were coded as discrete or discontinous characters. The phylogenetic trees (or cladograms) were calculated based on a matrix based on TFBS information.

2.2.19.1 Phylogenetic analysis of VNTR/ECR based on putative TFBS using Alibaba 2.1

The VNTR/ECR sequences were analysed with the Alibaba 2.1 software (http://www.gene-regulation.com/pub/programs/alibaba2/index.html, Grabe 2002) to

identify putative TFBS. Alibaba 2.1 detected binding sites for TFs (TFBS) (Figure 2.5a) based on TRANSFAC 4.0 database. To identify TFBS in VNTR and ECR sequences I used used the following parameters: minimum match conservation: 75%; min number of sites 4 conserved in score of pair similarity to known sites 50, matrix width in bp 10. Using these parameters the software generated a list of putative binding sites for each VNTR/ECR (for example see Figure 2.5b). This method was developed as recombination processes produce variation at the VNTRs loci. This recombination makes difficult to aligned VNTR sequences. As alignments are commonly used for the study of the evolution of DNA sequences, alignment of VNTRs introduce great number of penalties in the calculation of phylogenetic trees based on DNA sequences alignment, making the calculation innacurate. The identification of a putative TFBS in a VNTR or ECR sequence was not assumed as probe of the interaction between the TF and the regulatory domains, but simple facilitated the analysis of repetitive and potentially functional sequences.

For the phylogenetic analysis the presence of each TFBS in the VNTR or ECR was considered as an independent evolutionary event and thus quantified as independent characters (e.g. USF= character 1, YB-1=character 2; Figure 2.5c). The absence of a TF in a VNTR sequence was considered a secondary loss and was coded as a new state of character. For example, if a TFBS for USF is present in the STin2 VNTRs of *H. sapiens*, this is coded as one state of character (e.g. coded as 0). If the TFBS for USF was absent from the STin2 VNTRs of *P. troglodytes,* this is coded as a different state of the same character (e.g. coded as 1). When one species presented more than one state of character (i.e. the TFBS was found both present and absent in the population) this was considered polymorphic in that species, and a new state of character was created. For example, if the TFBS for USF was absent [=0] and present

[=1] in the *Gorilla* sp. STin2 VNTRs, this character would be coded as state "2" for that species (Figure 2.5).

a ***C. aethiops* STin2 VNTR**

```
seq(    0..   59)     ccatggctgtgacctggggtgggctgtgaaccggggtgggctgacccccaggtgggctgt
Segments:
2.1.2.2    6    15          ==RXR-beta
2.3.1.0   15    24                ====Sp1===
2.3.1.0   30    41                      =====Sp1====
1.3.1.2   35    44                           ====USF===
2.3.1.0   40    49                                ====Sp1===
2.3.1.0   47    57                                     =====Sp1===
2.1.1.4   56    65                                          ====
2.1.2.1   56    65                                          =RAR
2.1.2.3   56    65                                          =T3R
9.9.721   56    65                                          ==RA
================================================================================
seq(   60..  119)     gacccgggtgggctgcaacctggggtgggctgt
Segments:
2.1.1.4   56    65    =ER===
2.1.2.1   56    65    -alph=
2.1.2.3   56    65    -alph=
9.9.721   56    65    R-beta
2.3.1.0   65    74         ====Sp1===
9.9.539   70    79              ====NF-1==
2.3.1.0   79    90                   =====Sp1====

13 segments in this sequence identified as potential binding sites
```

b

```
1. SP1O= Present; 1= Absent
2. USF O= Present; 1= Absent
3. ER O= Present; 1= Absent
4. RAR-α O= Present; 1= Absent
5. T3R-α O= Present; 1= Absent
6. RAR-β O= Present; 1= Absent
7. NF1 O= Present; 1= Absent
8. RAP1 O= Present; 1= Absent
9. Ttx O= Present; 1= Absent
10.RxR-β O= Present; 1= Absent
11.CPE-bind O= Present; 1= Absent
```

c

```
                12345678901234
C. aethiops     00000001101000
M.nigra         01000110000000
```

Figure 2.5 Example of matrix building based on potential TFBS found in the VNTR/ECR. (a) The VNTR sequence (at the top of each section in red font) was analysed using the Alibaba 2.1 software. The potential TFBS identified by the Alibaba 2.1 (as indicated in blue font in Figure 2.5a) where displayed in an output file. (b) The TFBS data (presence or absence of a type of TFBS in the sequence) was transformed into a numerical format and considered as evolving units or "characters" (number 1-8 in Figure 2.5b). (c) The coded TFBS data was used to build a matrix of characters (in the top row are the number of each character which correspond to TFBS list in Figure 2.5b).

2.2.19.2 Calculating phylogenetic trees using Parsimony

Phylogenetic trees were built from a matrix of characters based on putative TFBS in the VNTR sequence (Figure 2.6). The phylogenetic trees were generated using the PARS program, which uses the general parsimony criteria, of the PHYLIP 3.6 package (Felsenstein, 1989). Briefly, the matrix of characters generated (see example in Figure 2.5c) was saved as a .txt file and uploaded into the PARS program. To calculate the tree, the program was set to use the most thorough search option and was set to search amongst the 100 best trees calculated. PARS calculated rooted trees where the hypothesised oldest species in the phylogeny was specified for the tree calculation. For example, when calculating the phylogenetic tree of the D4ECR1 of mammals (Figure 6.7), the ECR sequence of *C. familiaris* was used as the outgroup or root, as the group from which this species evolved is thought to have separated from the group of mammals that gave rise to primates earlier than rodents.

The default settings were chosen for all other options in the tree calculation. If one most parsimonious tree (amongst 1000 trees) was found, the length of the branches (BL) in the tree represented the evolutionary distance (expressed in number of steps) between the species. The program PARS produced two output files: outfile (A) which contained the phylogenetic tree (s) produced, and number of steps needed to produce the tree and a second file or outtree (B), which describes the phylogenetic tree in the parentheses notation. This second output file was used as the input file for the schematic representation of the phylogenies calculated (cladogram) which were obtained using the program DRAWGRAM.

a

```
Discrete character parsimony algorithm, version 3.64

One most parsimonious tree found:

              +Hsapiens
              |
  +-----------4     +-----Ppaniscus
  |           |     |
  |           +-----2-----Ggorilla
  |                 |
  |                 |     +Ptroglodyt
  |                 +-----3
  |                       +-----------------------------Ppygmaeus
  |
  1-Mmulatta
  |
  +Tgelada


requires a total of     11.000

  between       and        length
  -------       ---        ------
     1           4          2.00
     4       Hsapiens       0.00
     4           2          1.00
     2       Ppaniscus      1.00
     2       Ggorilla       1.00
     2           3          1.00
     3       Ptroglodyt     0.00
     3       Ppygmaeus      5.00
     1       Mmulatta       0.00
     1       Tgelada        0.00
```

b

```
((Hsapiens:0.00,(Ppaniscus:1.00,Ggorilla:1.00,(Ptroglodyt:0.00,Ppygmaeus:5.00)
:1.00):1.00):2.00,Mmulatta:0.00,Tgelada:0.00);
```

Figure 2.6. Calculating phylogenetic trees using PARS. The software calculated a phylogenetic tree based on a TFBS matrix (e.g. the produced in Figure 2.5c) and this analysis produced an outfile (2.6a) and an outtree (2.6b). These output files were used to create a phylogenetic tree using the program DRAWGRAM (see section 2.1.19.3).

2.2.19.3 DRAWGRAM

DRAWGRAM (from the PHYLIP package) was used to create graphic representations of the phylogenies calculated using PARS. Briefly, the outtree file produced by PARS was uploaded into the program; using the setting options of the program the font type, tree style and root of the tree were specified. The default settings were chosen for all other options of the program. The software produced a preview tree (Figure 2.7a) and after verifying the correct settings have been selected, a plot file (cdr) was created. The plot file was accessed in Corel draw X3 (Figure 2.7b) were the final editing of the tree (changing font size/style, placing on the branches full scientific names of the taxa, etc) was done.

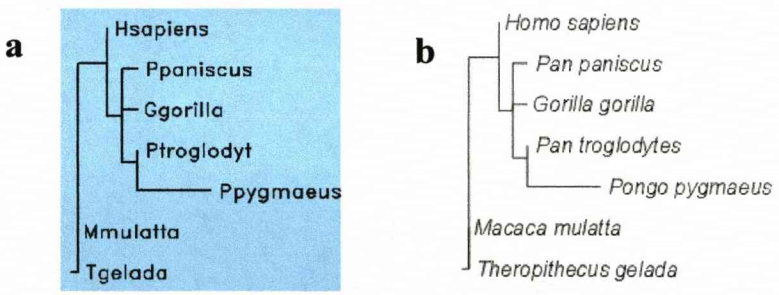


Figure 2.7 Graphic representation of the calculated trees using DRAWGRAM. The program creates a preview tree using outtree generated by PARS (a). After preview, a plotfile is created and this file is opened in Coreldraw 3X were final editing takes place (b).

2.2.20 Bioinformatics

2.2.20.1 BLAT

Sequences of plasmids and PCR products were compared to human genomic sequences using BLAT (Kent, 2000) from the UCSC browser. Briefly, sequences were uploaded in fasta format and a search was executed. Matching sequences in the human genome found by BLAT were displayed in a list of links. The top hit link was browsed to visualize the match with the human genome. To analyse the similarities between the human genome and the uploaded or query sequence, the bar representing the query sequence aligned to the human genome sequence was clicked, this produced a new link that displayed a letter-by-letter comparison between the two sequences. Identification of minisatellites in the human (or any other genome available in this browser) was possible by selecting the "full" option of the simple microsatellites tool of the browser. In addition, the conservation tool of BLAT created multiple sequence alignments, including genomes of species representing major vertebrate groups. Alignments of the most conserved regions around the VNTR were used to design primers to amplify the region in all primates (Figure 2.1). To identify ECRs located outside coding and promoter regions, entire gene loci were screened by setting on the conservation and regulatory potential tools of the BLAT browser of the UCSC webpage.

2.2.20.1 ECR brower

Similarly to BLAT, the ECR browser was used to identify non-coding ECR regions with putative regulatory potential (Figure 2.8). A gene name (following Genecards nomenclature) was loaded in the program and a list with links presenting close hits was displayed. The correct link was browsed and the program produced a

graphic comparison between genomes of species representing major vertebrate groups (human, macaque, rat, mouse opossum, chicken, toad and fish). This comparison allowed identification of evolutionary conserved sequences. Similarity between non-human vertebrate sequences and the human sequence was represented by vertical coloured lines which form coloured blocks (flanking intergenic= red; intronic=pink; exonic=blue, UTR=yellow; minisattelite=green). The program was set to identify non-coding ECRs that shared 70%-85% similarity between other vertebrate and human genomic sequence. When a block was present in distantly related mammalian species or other non-mamalian vertebrates, the sequence was considered as a non-coding ECR.

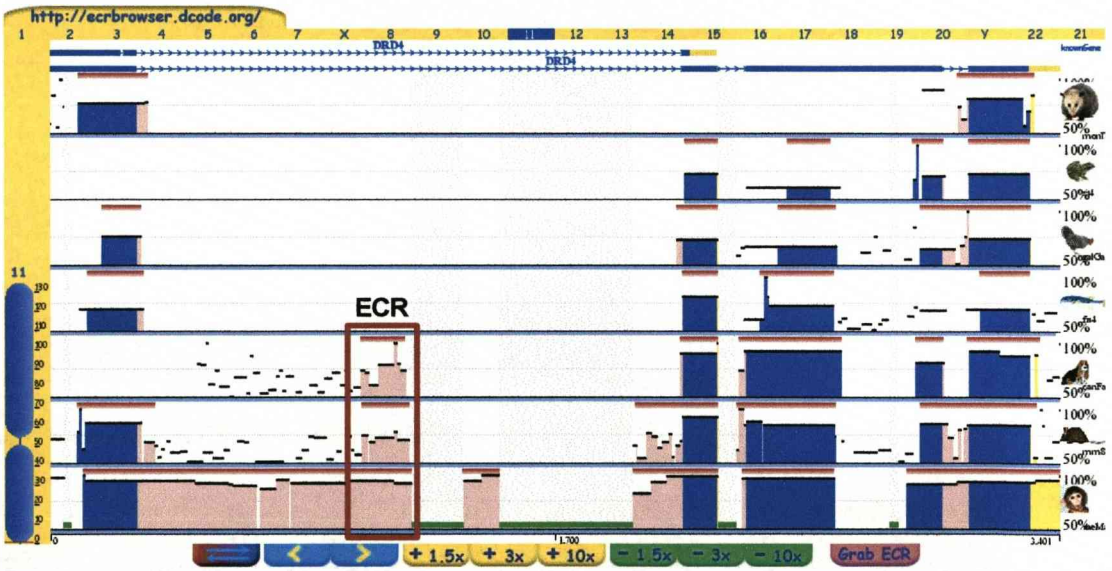


Figure 2.8 Identification of non-coding ECR using the ECR browser. Gene name (e.g. DRD4) was uploaded in ECR browser and the program created a vertebrate genome comparison. Similarity between the human and other vertebrates sequences was indicated by coloured bars (blue = coding, pink = intronic). Here, an intronic sequence that present great similarity across eutherian mammal sequences is circled in red, and was considered as an ECR within the DRD4 gene.

Chapter 3 Evolution and diversity of the VNTRs sequences of the SLC6A4 (promoter and STin2 VNTR) and DRD4 (exon 3 VNTR) genes

3.1 Introduction

Several studies have shown that the modern human population exhibits great sequence and copy number variation in the VNTRs of the SLC6A4 and DRD4 genes. Moreover, when the homologous VNTRs of non-human primates were studied, it was clear that the different species present VNTRs that differ in length and primary sequence (Lesch et al., 1997, Inoue Murayama et al., 2003; Livak et al., 1995). The transcriptional properties of these human and some of the non-human primate VNTRs have been demonstrated *in vitro* and *in vivo* (Miller et al., 2001, Inoue-Muruyama et al., 2003). These studies suggested that variation in the sequence of VNTRs might have implications in the way neurotransmission regulation varies amongst primates. Consequently, the study of the VNTR variation may have relevance for understanding the evolution of the regulation of modern human behaviour and cognition. The findings raise the following questions:

(1) When during evolution did these sequences became repetitive and acquire regulatory properties,

(2) How have these VNTRs evolved to their modern form

(3) Is there a correlation between the behavioural/cognitive differences across primates and the variation in these VNTR sequences?

In spite of the relevance of these questions, none of these has been fully explored. Therefore, in this chapter I investigated the diversity of selected VNTRs located in the DRD4 and SLC6A4 genes and the evolution of these VNTRs in primates using a phylogenetic approach.

3.2 Aims

The first aim of this chapter was to investigate the origin of selected VNTRs in the SLC6A4 and DRD4 genes. For this study, I analysed sequences of homologous regions obtained from different vertebrates obtained by PCR amplification of DNA samples listed in Table 2.2 (obtained for this study) and from the NCBI database. The second aim was to extend the knowledge on the diversity of the copy number and sequence of the VNTRs present in the DRD4 (exon 3) and in the SLC6A4 (promoter and intron 2) genes from different primates. In order to do this the different VNTR regions were amplified by PCR and sequenced. The third aim was to investigate the evolution of selected VNTRs, with emphasis on identifying the potential functional differences emerged during evolution. To do this I conducted phylogenetic reconstructions based on the TFBS found in the VNTRs identified by the Alibaba 2.1 program.

3.3 Results

3.3.1 The SLC6A4 gene

Studies have revealed the presence of two regulatory VNTR domains (STin2 and promoter VNTRs) in the serotonin transporter gene (SLC6A4) of primates. These elements are suggested to play an active and key role in the regulation of the SLC6A4 gene expression, and have been correlated to differences in serotonin availability in the CNS and serotonin related behaviour of modern humans *H. sapiens* and the rhesus macaques (*M. mulatta,* family Cercopithecidae) (Barr et al., 2004; Hranilovic et al., 2004). The absence of these VNTRs in the rodent (from both *M. musculus* and *R. norvegicus*) SLC6A4 genes (Heils et al., 1998) has been interpreted as that the VNTRs in the SLC6A4 gene are exclusive to primates. However, recent studies show that VNTR elements in other neurotrasmitter genes (DRD4 gene) are present in other

animal, such as *Canis familiaris* (Inoue-Murayama et al., 2002b). To determine whether these two VNTRs are present in the SLC6A4 genes of other vertebrates or mammals, I investigated this sequence in serotonin transporter homologous loci.

3.3.1.1 Origin of the SLC6A4 STin2 VNTR

To investigate the presence of the STin2 VNTR in the genomes of mammals, I amplified this intronic region by PCR from DNA samples of species representing different mammal groups. This included members of the order Chiroptera "bats" (microchiroptera: *C. perspicillata* n=1, and macrochiroptera *P. rodricencis* n=1), Rodentia (rodents, *M. musculus* n=1) and primates (apes, monkeys and prosimians, specimens listed in Table 2.2). The primers used for PCR were designed to prime conserved sequences across mammals (defined using the UCSC BLAT conservation tool) flanking the VNTR (see appendix 1).

The results of the PCR amplification are presented in Figure 3.1. In brief, the PCR reaction amplified intronic regions from DNA samples of higher primates, included in the families Cercopithecidae (old world monkeys) and Hominidae (great apes and modern humans). The sequences of the primers used were conserved across *H. sapiens* and *M. musculus* DNA, however the PCR was unable to amplify a product from the DNA samples of rodents. Although this negative PCR result does not discard the presence of a VNTR in the SLC6A4 genes of rodents, this does confirm previous findings by Heils (1998), which previously investigated the promoter region of *Mus musculus* SLC6A4 gene and did not identify a VNTR like sequence.

Therefore, I searched for a VNTR like sequences in loci homologous to the intron 2 in different mammal sequences, obtained using the UCSC BLAT browser (conservation tool). The browser produced a conservation alignment formed by the mammalian sequences homologous to the *H. sapiens* STin2 VNTR locus (Figure

3.2a). This included sequences of *M. mulatta*, *M. musculus*, *Canis familiaris* [domestic dog] and *Equus caballus* [horse]). My alignment showed that an 18 bases long conserved sequence **AcAGaCCAcCCCtgGGTC** (capitals represent conserved nucleotides and lower case indicate variable nucleotides in Figure 3.2a) was also present in the second intron of the SLC6A4 genes of dog, horse and mouse genes (highlighted in red in Figure 3.2a). However, in the sequences of *M. mulatta*, and in *H. sapiens* this 18 bp long sequence is interrupted at nucleotide 12, from were the repeats seem to have expanded. The sequences of *M. mulatta* and *H. sapiens* are arranged in tandem and indicated by roman numbers (I-XII for the *H. sapiens* VNTR).

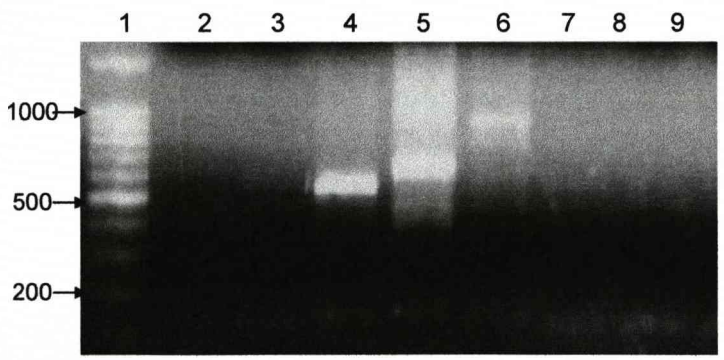


Figure 3.1 PCR of the STin2 VNTR locus from DNA of primates, rodents and bats. A VNTR was amplified from DNA of higher primates (e.g. *H. sapiens*, *P. troglodytes* and *Gorilla* sp) and not from DNA of rodents or bats. Lane 1= 1kb DNA ladder; 2=*R. norvegicus;* 3=*M. musculus;* 4=*H. sapiens;* 5=*P. troglodytes;* 6=*Gorilla* sp.; 7=*P. rodricensis;* 8=*C. perpicillata* and 9= negative control.

I also retrieved homologous sequences (to the STin2 VNTR) found in the SLC6A4 gene in two primitive mammals: *Ornithorhynchus anatinus* (platypus, Subclass Monotremata) and in *Monodelphis domestica* (common opossum, order Marsupialia, Subclass Metatheria). In *O. anatinus* I found a 14 bp sequence that shows homology to the 18 base pair sequence found in other mammals (highlighted in

red in Figure 3.2b). In this species, the conserved sequence is interrupted by a 146 bases long sequence. This 146 bp sequence possesses similar motifs to the found in the sequence of the primate VNTR (highlighted in blue, Figure 3.2b), but these motifs are not arranged in tandem. In the SLC6A4 gene of *Monodelphis domestica* I was unable to identify any homologue to the 18 bp sequence.

The results suggest that the VNTR is not found in canids or in ungulates like *E. caballus* and *C. familiaris*, and confirms its absence from the *M. musculus* and genes. Nevertheless, in *O. anatinus* the presence of a 146 sequence contained between the conserved 18 bp sequence which contain similar motifs to the found in the VNTRs of primates suggest that, this portion may have evolved by parallel evolution.

a

b

Platypus (*Ornythorhynchus anatinus*)

atctcccaatcccgctccgggcccctctccaatccccgcgcgctccggccgatcccgctctccaccagc
ccgttccgcgggcttcccggcgcccttcccggggcccgaccacccctcgccccacccgttagcag
cctggactccctaccgggtgggggtc

Figure 3.2 Origin of the STin2 VNTR in mammals (a) Alignment of the intron 2 sequences homologous to the STin2 VNTR in primates Sequences from modern human, rhesus macaque, lab mouse, domestic dog and horse were aligned manually. In this alignment, all non-primate mammals have of one copy of this repeat. The repeats forming the 12 repeats VNTR (n *H. sapiens*) are highlighted using either bold font or by underlining and are numbered with roman symbols. The sequences in red font (at the beginning and end of the alignment in 3.2a and 3.2b) indicate the single repeat unit found in other mammals (separated for better alignment). **(b) Shared motifs between primates and primitive mammals.** Some motifs were found in the VNTRs sequences of primates (human and rhesus macaque sequences, Figure 3.2a) and the platypus (Figure 3.2b). These motifs identified by eye and are highlighted in blue. In 3.2b, the portion of sequence believed to share homology with the 18 bp sequence of 3.2a (described in page 91) was highlighted in red font.

3.3.1.2 Origin of the 5' promoter VNTR in the SLC6A4 gene

I investigated the presence of a VNTR in the 5' promoter region of the SLC6A4 gene of mammals and vertebrates to determine the origin of this domain. I amplified the 5' promoter region from primate and rodent species by PCR (Figure 3.3) and analysed the homologous region in the serotonin transporter gene of vertebrates available in the UCSC genome browser.

The PCR amplification confirmed previous study by Heils (1998), which amplified a product containg the promoter VNTR from DNA of higher primates (modern humans, apes and old world monkeys), but not from DNA of lower primates (new world monkeys) nor from the DNA of *M. musculus* or *R. norvegicus*. However, examination of the 5' promoter region of the serotonin transporter gene in other vertebrates demonstrated that a tandemly repetitive sequence is not unique to the SLC6A4 gene of higher primates. I obtained an alignment (Figure 3.4) formed by sequences located at a homologous loci in the 5' promoter of their SLC6A4 genes of *Gallus gallus* (chicken), *O. anatinus* (platypus), *Gasterosteus aculeatus* (three spine stickle back fish) and of *Xenopus tropicalis* (pipid frog).

Analysis of these non-mammalian sequences showed that these sequences not only share great homology with the primate 5'VNTR (Figure 3.4) but are also arranged as tandem repeats (TR) (Figure 3.5). Although there is great homology between the primate VNTR sequences and the non-mammalian TR sequences there were differences in their organization. For example, whereas the primate VNTR is arranged in repeat units of 20 to 23 nucleotides, the TR of *G. gallus* presented 25 repeat units formed by 11 bases. Similarly, the TR of *O. anatinus* presented 13 repeat units formed by 23 bp; the *G. aculeatus* presented a TR with 96 bp and in *X. tropicalis* the TR was formed by 11 repeats formed by 30-36 bp. Furthermore, the

primate VNTRs and these vertebrate TRs shared many motifs (examples are highlighted in grey, Figure 3.5). Based on these common motifs and the proximity of these elements to the 5' promoter, the findings suggest that these vertebrates may also present a 5' promoter VNTR acting as a regulatory domain modulating for the expression of the serotonin transporter gene. However, more individuals from each vertebrate studied must be analysed to confirm intra-specific variation of the TR sequences to clasify them as VNTRs.

The findings propose two explanations for the origin of the promoter VNTR in primates. The first explanation is that a repetitive sequence rich in GA content appeared in the 5' promoter of the SLC6A4 gene of vertebrates approximately 450 mya, time proposed to have passed since all major groups of vertebrate evolved (Meyer and Zardoya, 2003). Therefore, this domain would have been lost in the mammal lineage and evolved *de novo* in the genomes of higher primates by parallel evolution. This would have occurred around 35 mya, approximately the time old world monkeys and apes originated. The second explanation poses that, given the great similarities between the primate VNTRs and the vertebrate TRs (Figure 3.5) the VNTRs in the 5' promoter region of the SLC6A4 gene of some mammalian species such as *M. musculus* and *R. norvegicus* has been lost; but persisted in others such as higher primates. As the regulation of serotonin related behaviours (i.e. aggression, dominance, sexual arousal, stress response) in very important in all vertebrates. It is possible that such an element exists in the 5' region of the SLC6A4 gene of other vertebrates. Furthermore, it is possible that such element exist in the SLC6A4 gene of other mammalian species not studied here or in previous analyses. Therefore, based on the similarities between the primate and vertebrate repetitive sequences, I proposed

that the VNTR in the 5' promoter region and the TR found in other vertebrates are homologous in origin.

Figure 3.3 Gel of PCR of the promoter VNTR region in mammals. The PCR amplified a VNTR region from DNA samples of higher primates but not from DNA of bats, rodents or lower primates. Lane 1 shows DNA ladder (1kb). Lane 2= negative control; lane 3=*Gorilla* sp.; lane 4=*P. pygmeaus*; lane 5= *Hylobates* sp.; lane 6= *M. sphinx*; lane 7= *C. aethiops*; lane 8= *L. catta*; lane 9=*M. musculus* and lane 10=*C. perspicillata*.

Figure 3.4 Alignment of the repetitive sequences found in the 5' promoter region of the SLC6A4 genes of different vertebrates homologous to the promoter VNTR of primates. The alignment was generated using the conservation tool of the BLAT search engine (of the UCSC genome browser). Great protions of the sequence are highly conserved throughout the evolution of vertebrates.

H. sapiens

aatgctggaggggtgcaggggg
gatgccgggggtgcatggggg
Gatgctgggggtgcaggggg
gatactgcgaggggtgcaggggg
gataatgggggttgcagggga
gatcctgggagaggtgcagggg
gatgctggaagggctgcaggggg
gatgctgggggtgcagggga
gatgctgggggggctgcaggggg
gatgctgggggtgcaggggg
gatgccgcgaggggtgaaggggggg
gataatgggggatgcaggagc
gatcctaggaggggtacagg
gagttgctgggaggttgcaggggg
gatgctgggagggctgcagt

O. anatinus

ggggttgccccgggggccggagg
ggggctgctcaggggggccggagg
gggctgctcaggggccggaggg
ggggctgttcagggtgccggagg
ggggctgccccgggtgccggagg
ggggatgctccgggggccggagg
ggggctgctcaggggccggagg
ggggatgctcaggggccagagg
ggggctgttcagggtgccggagg
ggggctgccccgggggccggagg
ggggctgctccgggtgctgcaag
ggggctgttcaggggccatagg
ggggctgctccgggtgccggagg

X. tropicalis

ggacgctgaaggatacttgggggggagccctggggg
aagctgaaaggatacttgggggggg---ccctggggg
aagctgaaggatacttgggg-agg--ccctggggg
acgctgaaggatacttgggggggg--ccctggggg
aagctgaaggatactt-gggggggg--ccctggggg
aagctgaaggatacttggggag-g--ccctagggg
atgcggaaggatacttgggggggag--ccctggggg
aacctgaaggatacttgggggg----ccctggggg
aagctgaaggatacttgggggggga---cct-gggg
gagctgaaggatacttgggggggggggccctggggg
aagctgaaggatacttgggggggg---ccctgg

G. gallus

atagatgctaa
gaatgattgag
ggatgctgctg
gtatgatactg
ggtgaataacg
ggaggatgctg
agaggatgcca
ggatgattctg
gaaagatgttg
gaaggatgcaa
gtaggacactg
ggagatgat
aagagaagctg
agaggacatt
gggaaatgctg
agaataagctg
ggaagatgctg
ggagaatgtag
gaaggatgctg
-agaaatactg
gaatagaaata
caaggaggatatta
gagtgatgctg
ggagaatgctg
aaaggatgctg

G. aculeatus

gatgctgggagggaaccaggtggatgctgggagggaaccagctgggagggaaccaggaggatgttgggagg
gatgctgggagggaaccaggtggatgctgggagggaaccagctgggagggaaccaggaggatgttgggagggaaccagctgtgag
ggaaccaggag
gatgttgggagggaaccaggaggatgctgggagggaaccagctgggagggaaccaggaggatgttgggagggaatcaggtgggag
ggaaccaggag

Figure. 3.5 Tandem repeats (TRs) found in the 5' promoter region of the serotonin transporter gene of *H. sapiens* and other non-mammalian vertebrates. The TRs sequences of vertebrates share motifs with the VNTR sequence of *H. sapiens* VNTR (examples of similar motifs are highlighted in grey). Although highly homologous, the TRs of these vertebrates are organised different from the *H. sapiens* VNTR.

3.3.1.3 Diversity of the copy number of the STin2 VNTRs in hominids.

The STin2 VNTR copy number variants (with 9, 10 and 12 repeat units) commonly found in modern humans exhibit differential transcriptional properties (Mackenzie and Quinn 1999; Klenova et al., 2004; Lovejoy et al., 2003). These differences have been propose to correlate with the variability of normal and pathologies of the emotional behaviour of modern human (Kremer et al., 2005; Mulder et al., 2005; Payton et al., 2005). The great apes and old world monkeys also present a polymorphic STin2 VNTR, which sequence and copy number varies across primates (Soeby et al., 2005). The variation in the STin2 VNTRs suggests that if functional *in vivo*, this VNTR may contribute to the diversification of serotonin related behaviour amongst different primate species. For this reason, I expanded on the current knowledge on the diversity of this VNTR by PCR amplification from samples of genera and species not studied before and increased the sample size of those previously analysed.

The allelic diversity (copy number) of this STin2 VNTR found in hominids (great apes) and cercopithecids (old world monkeys) shown in Figure 3.6. In brief, amplification of this locus by PCR confirmed that the STin2 VNTR exhibit great diversity across hominids and cercopithecids. Furthermore, the present study demonstrates that the diversity of copy number is higher in hominids than in cercopithecids. Indeed, whereas the VNTRs of hominids present from 6 to 40 repeat units, all cercopithecids VNTRs amplified present 5 repeats.

In hominids, *P. troglodytes* (common chimpanzee) exhibited four different copy number variants (with 18, 19, 21 and 23 repeat units per VNTR). In this study 32 samples were amplified, with most individuals presenting a VNTR with 19 repeat units (n=18), followed by those carrying a VNTR with 18 (n=10), 21 (n=2) and 23

repeat units (n=2). In summary, the present study has identified 2 more copy number variants to the VNTRs previously reported by Soeby (2005) and those deposited by Inoue-Murayama in the NCBI database (VNTRs with 19, 23, 24 and 25 repeat units). It is noteworthy that the samples studied were obtained from a colony housed in Chester zoo (Table 2.2), which may be subject to inbreeding, affecting the genetic diversity of the colony. However, as previous studies found VNTRs with similar number of repeat units in unrelated individuals of *P. troglodytes* this is indicative that the variation found within this *species* housed in Chester Zoo is representative of the variation found in populations in the wild.

In the *Gorilla* sp., I was able to amplify the VNTR region from one out of three specimens. This individual possessed a VNTR with 40 repeat units, which was the same number of repeat units previously reported for this species (Soeby et al., 2005). In *P. pygmaeus* (orang-utan), two out of three DNA samples were amplified by PCR. This study constitutes the first report of the STin2 VNTR sequence of *P. pygmaeus* and shows that its VNTR was formed by fewer units (6 repeats) than the other hominid VNTRs, and that it is similar in length to the cercopithecids VNTRs (with 5 repeat units) (Soeby et al., 2005).

The present study suggests that the STin2 VNTRs of *P. troglodytes* presents more copy number variants (7) than *H. sapiens* (5). This result agrees with previous studies, which showed than indeed, within the small population of *P. troglodytes* (100000 to 200000 individuals; Gagneux, 2002), there is greater genetic variability than the found in the entire human population (6.5 billion, united nation website: http://esa.un.org/unpp/). This great diversity in *P. troglodytes* populations may be correlated to the geographic isolation of its populations in the wild. Such geographic isolation seems to also favoured the expansion of rare alleles in isolated populations

of modern humans such as the mbuti tribe, which are the only modern humans that exhibit a STin2 VNTR with 11 repeat units (Gelernter et al., 1999).

In summary, the present study demonstrates that great variation in the STin2 VNTR has been accumulated during the speciation of hominids (Figure 3.6). It can be deduced from the results that the ancestral hominid may have had a VNTR with 5 or 6 repeat units, similar to the found in *P. pygmaeus* and cercopithecids. This finding suggest that the STin2 VNTRs in *H. sapiens* would have either originated by an expansion from the original ancestral hominid VNTR or by a reduction (in repeat units) from the VNTR present in the last ancestor shared with *P. troglodytes* and *Gorilla* sp. To determine this, it is necessary to analyse the evolution of the hominids and cercopithecids STin2 VNTR sequence using a phylogenetic approach.

3.3.1.4 Diversity of old world monkeys STin2 VNTR copy number

The diversity of the STin2 VNTR copy number of cercopithecids was investigated in two species of cercopithecids (old world monkeys) not studied before (*Mandrillus sphinx* n=1, and *Cercopithecus aethiops* n=1; Figure 3.6). The PCR amplification demonstrated that these species each presented five repeat units per VNTR. These findings confirm previous research conducted by Soeby (2005) which also found STin2 VNTRs with 5 repeats in cercopithecids. The present study also suggests that although comparatively similar time have passed between the different cercopithecid and hominid species analysed (Harris, 2000), there is relatively less variation in copy number in the STin2 VNTRs in cercopithecids than in hominids. This result suggests that different forces may be driving the evolution of the STin2 VNTR sequence in both primate families, producing great diversity in VNTR copy

number variants in hominids and maintaining low variability in the VNTRs of cercopithecids. However greater sample sizes are needed.

Furthermore, this study sheds light on the origin of the hominids VNTR. The number of repeats forming the VNTR of *P. pygmaeus*, the earliest hominid and that of all cercopithecids studied to date is similar (5 and 6 repeat units per VNTR respectively; Figure 3.6). As mentioned in the previous section, the length of the STin2 VNTRs of *P. pygmaeus* suggested that the ancestral hominid would have also had a VNTR with few repeat units (5-6). The comparison of the *P. pygmaeus* VNTR to the VNTR of cercopithecids, a family of primates often used as to represent a "primitive" hominid state, offers support to this theory. However, further analysis on the diversity of the primary sequence of this VNTR is required to determine its most likely evolutionary pathway.

Diversity of the STin2 VNTR of the SLC6A4 gene of primates

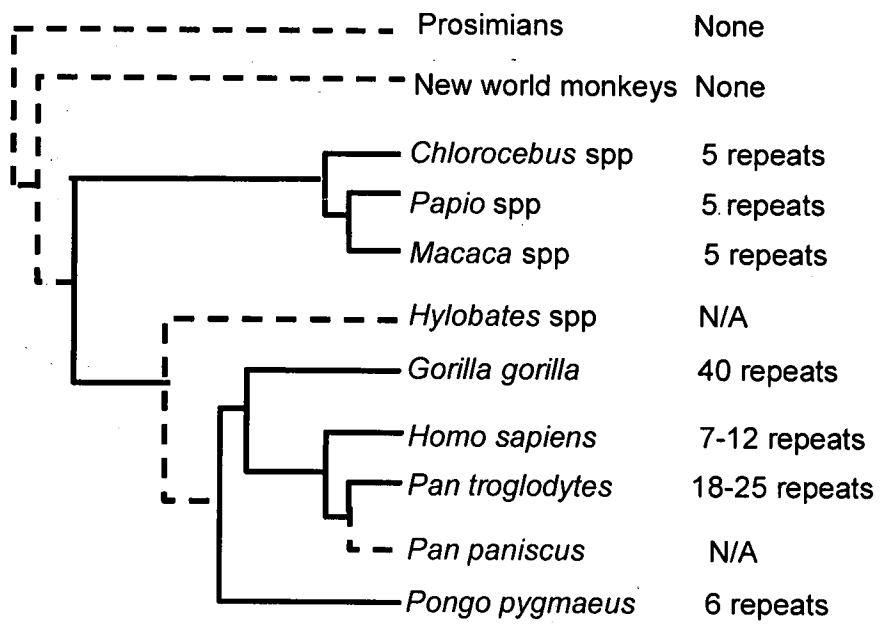


Figure 3.6 The diversity of the STin2 VNTR copy number in primates. The primate species are positioned in a phylogenetic tree, which represents their most commonly accepted evolutionary relationships (Based on Page and Goodman, 2001). In the tree, black solid lines indicate that in those branches, the presence of a VNTR has been confirmed. Dashed lines show primates where the presence of this VNTR has not been determined or where the VNTR is not present. The hominid VNTRs show greater diversity in copy number (6-40) than that of old world monkeys (with 5 repeats).

3.3.1.5 The diversity of the primate STin2 VNTR sequence.

The change of even one base can affect the transcriptional activity of a regulatory domain. In spite of the relevance of the study of the STin2 VNTR sequence changes for understanding the transcriptional properties of these polymorphic domains, few studies have taken in consideration this diversity. I therefore investigated the diversity of the STin2 VNTRs sequences in primates using primary DNA sequences obtained by PCR amplification and from the NCBI database. To simplify sequence information I named each repeat unit found in the different VNTRs based on nomenclature presented by Lovejoy (2003). The sequences arranged in repeat units are detailed in appendix 3.

3.3.1.5.1 Alignment of the hominid STin2 VNTR repeat units

I aligned the STin2 VNTR sequences of hominids to visualize the diversity of the VNTR repeat units, where these have been conserved in the VNTRs and where changes in number of repeats had occurred to cause the observed differences in copy number amongst hominids. The repeat units VNTR were manually aligned for best fit. The alignment is shown in Figure 3.7.

3.3.1.5.1.1 The hominid STin2 VNTRs present primary sequence variation and the presence of species-specific repeats

The alignment of the sequences revealed primary sequence variation in of the the STin2 VNTRs of hominids. For example in *P. troglodytes,* from 35 DNA samples amplified by PCR, I found 18 different VNTR sequence variants. The VNTR with 19 and 18 repeats both presented 7 variants and the VNTRs with 21 and 23 repeats both presented 2 different variants (Figure 3.7). In total, this study increased the number of unique repeat units reported for *P. troglodytes* from 7 (Soeby et al., 2005) to 25. The *Gorilla* sp. VNTR sequence amplified presented 40 repeat units. This sequence

differed from the one previously published by Soeby (2005) in 11 repeats. In *P. pygmaeus*, the two individuals VNTRs presented the same sequence. Considering that the STin2 copy number variation data used in this study is based on analysis of many more *H. sapiens* individuals than of *P. troglodytes* individuals (n>1000 and n= 27 respectively); the results presented here confirm that *P. troglodytes* is genetically more diverse than *H. sapiens* at the STin2 VNTR locus.

The sequence analysis also demonstrated the presence of a species-specific repeats in *H. sapiens*, *P. troglodytes* and *Gorilla* sp. For example, all *H. sapiens* variants present "e" and "b" repeats not found in other hominids. Similarly, the *P. troglodytes* VNTRs present 65 unique repeats respectively. In addition, in this species I identified a repeat unit formed by 15 bases instead of the typical 16 or 17 bases (repeat "z"). In *Gorilla* sp., I found 9 unique repeats. Conversely, the *P. pygmaeus* STin2 VNTR did not show species-specific repeats, as it was formed by three sets of "dg" repeats found in other hominid STin2VNTRs.

3.3.1.5.1.2 The presence and position of a functionally relevant repeat is variable amongst hominids STin2 VNTR

Our group has demonstrated that oligonucleotides formed by one or two repeat units of the STin2 VNTR, or oligonuclotides that expand from the second half of one repeat to the first half of the following repeat unit (obtained from the human STin2 VNTR) are capable of supporting transcriptional activity *in vitro* (Lovejoy et al., 2003). Amongst them, one of the oligonuclotides that showed the greatest transcriptional activity *in vitro* contained the "f" repeat (Figure 3.7). The 3 most commonly found variants of the human STin2 VNTR (with 9, 10 and 12 repeat units) differ in the number of "f" repeats (one "f" repeat for the 9 VNTR and two for the 10 and 12 VNTRs). Furthermore, the 9 repeats VNTR is a rare allele in modern human

populations and has been associated with the onset of behavioural disorders (Battersby et al., 1996). Thus, it is possible that the lack of one "f" unit as seen in the 9 VNTR correlates with deficient regulation of SLC6A4 gene expression in modern humans.

I therefore analysed the presence, location and number of "f" repeats found in the STin2 VNTRs of hominids. This analysis showed that the "f" repeat is present in all *P. troglodytes* and *Gorilla* sp. analysed, however, it is absent from the VNTR of *P. pygmaeus* (marked by blue line rectangles, Figure 3.7). The absence of the "f" repeat from *P. pygmaeus* indicates that this has evolved recently in African great apes (*P. troglodytes* and *Gorilla* sp.) and modern humans.

There are differences in the number and position of this repeats in *H. sapiens*, *P. troglodytes* and *Gorilla* sp. (Figure 3.7). The "f" repeat is present three times in the *Gorilla* sp. VNTRs and once or twice in the *P. troglodytes* and *H. sapiens* VNTRs. In the latter two species, the "f" repeats are positioned in the centre and at 3' end of the VNTR. In *H. sapiens*, the presence of the "f" repeat is variable in the centre of the VNTR whilst in *P. troglodytes*; its presence is variable in the 3' end of the VNTR. All these differences in the position and number of the "f" repeat, amongst other repeats forming the hominid VNTRs suggest possible functional divergence amongst the regulatory properties of the STin2 VNTRs of hominids. Moreover, it is possible that the lack of "f" repeats in STin2 VNTRs variants of *P. troglodytes* could be associated to intra-specific behavioural differences in this species.

3.3.1.5.1.3 Inferences of the evolution of the hominid STin2 VNTRs based on their repeat units

The alignment of the STin2 repeat units helped understand how the different hominid VNTRs may have evolved. These VNTRs appeared to have evolved by

expansion and reduction of the specific repeat units, occuring at the 5' and 3' regions of the VNTRs. For example, the *Gorilla* sp. VNTR presents 13 repeats positioned at the 5' ("dgdgfdc" and "dgdδg"), absent from all other hominid VNTRs (marked by red rectangles in Figure 3.7). Similarly, the length difference between *P. troglodytes* and *H. sapiens* is mainly caused by the occurrence of repeats at both the 5' and 3 'ends (e.g. "xw" and "vgu" repeats) of the *P. troglodytes* VNTRs. The 3' regions of the STin2 VNTRs of *P. troglodytes* seem more prone to variation than the 5' end; consistently this area has more new repeats than other regions (marked by black solid lines in Figure 3.7).

The alignment demonstrated that a stretch of "d" and "g" repeats as those found forming the *P. pygmaeus* VNTR also occurs in other hominids VNTRs (indicated by an arrow head in Figure 3.7). For example, the *Gorilla* sp. VNTR presents two stretches of "dg" repeats, the first aligns with the *Pongo*'s "dg" repeats and a second stretch aligns with repeats at the VNTRs 3' end. In *P. troglodytes* I found two types of VNTR variants, those that exhibited a "dg" pair that aligned with the "dg" repeats found in *P. pygmaeus* (e.g. Pantro19c in the alignment) and those variants which only presented one "d" repeat aligning to the "dg" chain in *P. pygmaeus* (e.g. Pantro18g in the alignment). In the first type the "dg" pair was often followed by "ggvguf" repeats. In the second type, the "d" repeat was followed by repeats unique to *P. troglodytes* and even unique to the specific VNTR variant (Figure 3.7). In the *H. sapiens* VNTRs, differences in the number of "dg" repeats contribute to the formation of the 10 and 12 VNTR variants (Figure 3.7). In summary, variation in "d" and "g" repeats appear to have contributed to the formation of new variants and expansion of the STin2 VNTRs of hominids.

The alignment showed that *Gorilla* sp. and *P. troglodytes* VNTRs shared more repeats (13, marked by dashed black lines, Figure 3.7) than either shared with *H. sapiens*. These results would suggest that the STin2 VNTRs of *P. troglodytes* and *Gorilla* sp. shared a common ancestor, not shared by *H. sapiens*. As this contradicts the most accepted theories on the evolution of hominids, it is necessary to address this question using a phylogenetic analysis. This analysis (based on the TFBS found in the STin2 VNTRs) is detailed in section 3.3.1.6.

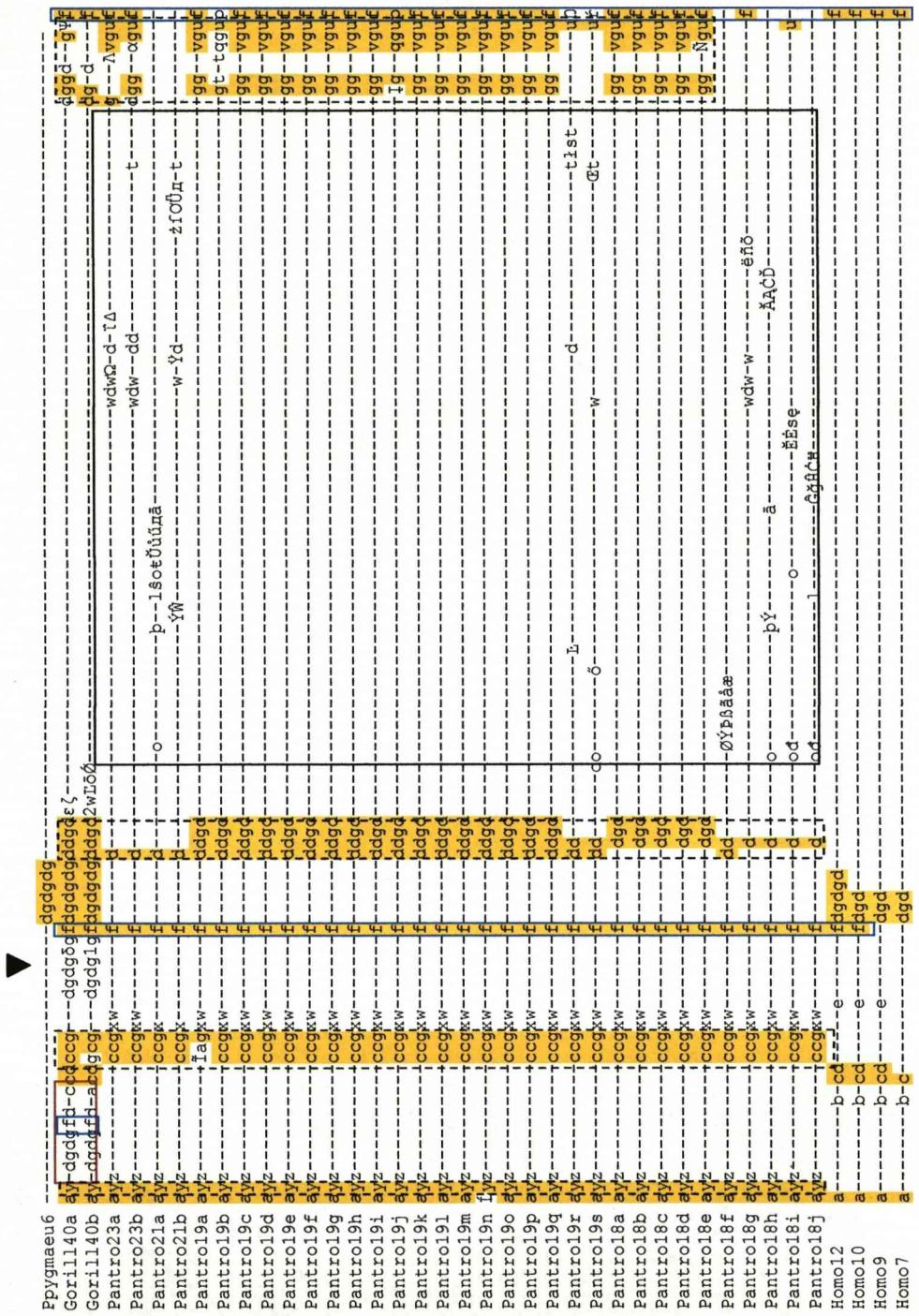


Figure 3.7 Alignment of the modern human and African ape STin2 VNTR repeats. In the alignment each symbol represents a repeat unit found in the STin2VNTRs of hominids, the sequences each symbol represent are detailed in annex 3.1. Intra-specific variation is predominantly present within the 3' end of the repeat. Repeats present in all species are higlighted in yellow and repeats present in one or two species are marked by dashed rectangles. Blue rectangles indicate the presence of "f" repeats. A red rectangle indicates the presence of unique gorilla repeats. The sequence of the STin2 VNTR with 11 repeats of *H. sapiens* is not available. Arrowhead indicate the alignment of the repeats found in *P. pygmaeus* to homologous repeat sections in other hominid STin2 VNTRs. Gaps introduced to optimise alignment.

3.3.1.5.1.4 The type and number of binding sites for YB-1 and CTCF varied in the primate STin2 VNTRs

The *H. sapiens* STin2 VNTR sequence contain Y-boxes (type a: gatgggct and g: ggtgggct, see Figure 3.8 for a sample) which bind the YB-1 protein. The presence of these boxes in the modern human VNTR sequences has been correlated to binding affinity for YB-1 (Klenova et al., 2004), and consequently differential regulation by YB-1. Furthermore, putative binding sites for CTCF have been suggested to mediate CTCF regulation of the STin2 VNTR activities (Roberts et al., 2007). The regulation of the VNTR activity via Y-boxes and CTCF binding sites is likely to play an important role in the regulation of SLC6A4 expression (section 1.12.1). Therefore, I investigated the presence of these Y-boxes and CTCF binding sites in the primate STin2 VNTR sequences. The results of this analysis are detailed in Table 3.1.

In brief, I found that the two types of Y-boxes are present in *P. troglodytes* and *Gorilla* sp, but are absent from *P. pygmaeus*. I found three a-Y boxes in the *Gorilla* sp. and one or two in *H. sapiens* and *P. troglodytes*. Considering that the specific Y-boxes in the *H. sapiens* STin2 VNTR can form different complexes with YB-1 *in vitro*, as demonstrated by EMSA (Klenova et al., 2004), the differences in Y-boxes found amongst hominid STin2 VNTRs could correlate with differential regulation of these VNTRs regulatory activities (Klenova et al., 2004). Furthermore, amongst these three species, there were differences in the distance between a-Y boxes within their VNTRs. This distance is greater in *Gorilla* sp. (13 or 17 repeats) and in *P. troglodytes* (8 repeats) than in *H. sapiens* (4 repeats). There were also differences in the number of g-Y boxes amongst the hominids VNTRs. This number was greater in *Gorilla* sp. (33-35), followed by *P. troglodytes* (12-13), *H. sapiens* (7-8) and *P. pygmaeus* (6) (Table 3.1).

The number of CTCF binding sites varied across primates STin2 VNTR as well based on binding sites defined by Klenova (2004). For example, the *Gorilla* sp. STin2.40 VNTR presented from 13 to 14 sites. In spite of the different number of repeat units, the number of CTCF binding sites in *H. sapiens, P. troglodytes* and *P. pygmaeus* overlapped. For example, *H. sapiens* exhibited between 3 to 4 sites for CTCF in its VNTRs whereas in the *P. pygmaeus* VNTR 3 sites were identified. The *P. troglodytes* exhibited from 1 to 6 binding sites for CTCF. The cercopithecids presented fewest number of CTCF sites, exhibiting typically 1 and occasionally 2 sites for CTCF binding (Table 3.1)

Since the distance, type and number of binding sites in these regulatory domains can affect their interactions with TFs, the findings offers further support to hypothesis that the sequence differences exhibit by the hominids STin2 VNTRs can correlate with functional divergence in their transcriptional properties mediated by YB-1 and CTCF (Roberts et al., 2007; Klenova et al., 2004).

	Binding sites for YB-1			Binding sites for CTCF
Species name	number of a-Y boxes	number of repeats separating a-Y boxes	number of g-Y boxes	Number of CTCF binding sites per VNTR
Macaca facicularis	0	NA	5	2
Papio papio	0	NA	5	2
Papio hymadryas	0	NA	4 or 5	1 or 2
Cercopithecus aethiops	0	NA	5	1
Mandrillus sphinx	0	NA	5	1
Pongo pygmaeus	0	NA	6	3
Gorilla sp.	3	13 and 17 repeats	33 or 35	13 or 14
Pan troglodytes	1 or 2	8	12 or 13	1 to 6
Homo sapiens	1 or 2	4	7 or 8	3 or 4

Table 3.1 YB-1 boxes (Y-boxes) in the STin2 VNTRs of primates. The number of repeats and spacing between them varies greatly amongst hominids

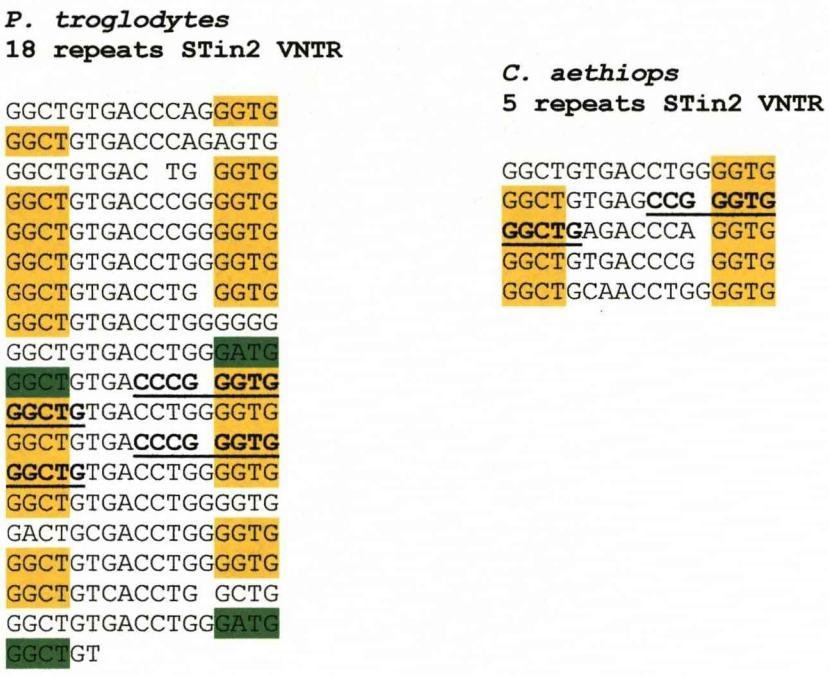


Figure 3.8 Y-boxes for binding to YB-1 and CTCF bindig sites in the hominid and cercopithecid VNTR sequences. Highlighted in yellow are g-Y boxes (in *M. sphinx* and *P. troglodytes*) and in green are a-Y boxes (only in *P. troglodytes*). Underlined nucleotides (highlighted in bold font) indicate potential binding site for CTCF (as defined by Roberts et al., 2007).

3.3.1.5.2 Diversity of the old world monkey STin2 VNTRs

The diversity of the STin2 VNTR of the old world monkeys has been linked to intra and inter-specific behaviour differences (Wendland et al., 2006b). Understanding the diversity of this STin2 VNTR can contribute to the understanding of the diversification of serotonin related behaviour mediated by the STin2 VNTRs in cercopithecids and importantly, the evolution of the hominid STin2 VNTRs. I therefore constructed an alignment of the repeat units found in all known cercopithecid STin2 VNTRs sequences to date (those amplified in previous sections and those from the literature). The repeats were named based on the nomenclature established by Lovejoy (2003) and aligned (Figure 3.9).

In brief, the STin2 VNTR sequences of *M. sphinx* and *C. aethiops* amplified in section 3.3.1.4 present similarities to previously published cercopithecids sequences (Soeby et al., 2005). However, some repeat units found in these species VNTRs are species specific. For example, *M. sphinx* presents "g", "d", "n" and "m" repeats (also found in other cercopithecid sequences) and one "c^5" repeat which is unique to this species VNTR. The VNTR of *C. aethiops* was formed by "g", "d" and "m" repeats and the species-specific repeats "c^3" and "c^4".

The variation of repeat units within the cercopithecids STin2 VNTRs occur at different sites (Figure 3.9). In three out of five VNTR sequences, this variation occurs at the 3rd repeat (from the right), however, variation was also observed in the 2nd (in *M. fascicularis*) and 5th repeats (in *C. aethiops*).

The analysis of the *C. aethiops* and *M. sphinx* VNTRs confirmed that all cercopithecids STin2 VNTR sequences studied to date bear one "d" and one "g" repeats. These two repeats occupy 1st and 4th positions in all sequences studies. The conservation of the position of these repeats and their presence across all primate

STin2 VNTRs (cercopithecids and hominids) suggests that these two repeats may have characterised the VNTR present in the primate that was the common ancestor to cercopithecids and hominids. Moreover, is possible that these "d" and "g" repeats play an important role in the *cis* regulatory function of these VNTRs. Interestingly, Soeby (2005) suggested that these are the most primitive repeats in the STin2 VNTR and could have given rise to all other repeats found in this STin2 VNTR.

To investigate the evolution of the cercopithecids STin2 VNTRs I compared their repeat variation and position. I ordered the species following a commonly accepted phylogeny of cercopithecids, based on molecular and morphological data (modified from Harris, 2000; Figure 3.9). This comparison demonstrated that the STin2 VNTRs of closely related species e.g. these contained in the genus *Papio* (*P. papio* and *P. hamadryas*) share more repeats than species more distantly related e.g. as *P. papio* and *C. aethiops*. This similarity suggests that the degree of divergence found in cercopithecids STin2 VNTR correlates to the evolutionary distance (and time passed) that exist amongst species.

This alignment also suggests that less variation in type and arrangements of repeats has occurred in the cercopithecid than in the hominid STin2 VNTRs. Indeed, there are 4 variants of the STin2 VNTR in cercopithecids and at least 37 in hominids. This finding suggest that evolutionary forces maintain low variation (purifying selection) at the STin2 VNTR locus which is greater in cercopithecids than in hominids; however, more specimens should be studied to confirm this hypothesis.

Finally, the similarity in the size of the cercopithecids *P. pygmaeus* VNTRs and the presence of "d" and "g" repeats, offers support to the hypothesis that the primitive hominid presented a STin2 VNTRs with fewer repeats that the found in *P.*

troglodytes and *Gorilla* sp. This consequently indicates that the STin2 VNTRs of hominids have undergone greater expansion than the cercopithecid VNTRs.

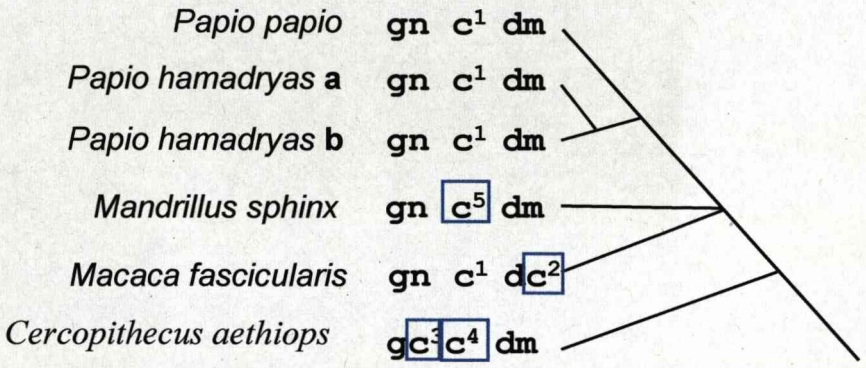


Figure 3.9 Diversity of the STin2 VNTR in members of the family Cercopithecidae. The repeats found in the STin2 VNTRs of old world monekys (Cercopithecidae) were represented by symbols, each was detailed in appendix 3.1. In this figure the species were arranged in a tree that represent their relatedness relationships (based on Harris, 2000). The position of the repeats that vary across STin2 VNTRs of these species was marked by blue lined-rectangles.

3.3.1.6 Evolution of the STin2 VNTR sequences

The alignment of the STin2 VNTRs of hominids and cercopithecids has offered some clues to their evolution; however, the evidence provided was not conclusive for the STin2 VNTRs of hominids and needs to be further investigated. I therefore reconstructed the evolution of the STin2 VNTR of primates by conducting a phylogenetic analysis based on the TFBS diversity in the VNTR sequences following protocol described in section 2.2.19.2. The identification of TFBS was not interpreted as evidence that these TF were necessarily interacting with the STin2 VNTR sequences. However, this novel method facilitated the processing of the information contained within the VNTR sequences for the phylogenetic analysis. The matrix produced based on the TFBS found in the VNTRs is shown in appendix 4.1. This analysis was conducted by the discrete parsimony program PARS, from the PHYLIP package (Felsestein, 1989).

The phylogenetic analysis of the STin2 VNTR matrix produced one most parsimonious tree (Figure 3.10a). In the cladogram, two VNTR sequences (of *C.s aethiops* and *M. fascicularis*) did not cluster with any others and remained as branches of "unresolved" phylogenetic ancestry. This occurs when the traits or characters presented by these sequences are not shared with other sequences included in analysis or when these features produce branching patterns that contradict each other. However, there is an internal branch formed by the VNTRs of the species of the genus *Papio* (*P. papio*, *P. sphinx* and *P. hamadryas*) and by the hominid STin2 VNTRs. This internal clade formed because these VNTRs share an evolutionary step,

involving changes in one TFBS (in this case, a site for USF; appendix 4.1), not shared by the VNTRs of *C. aethiops* and *M. fascicularis*.

The hominid sequences clustered in a separate branch because they shared one evolutionary step (in this case the presence of a binding site (BS) for *trithorax*, Ttx; appendix 4.1). In this hominid branch, the VNTR of *P. pygmaeus* occupies a basal position. This coincides with the known position of this species in most known phylogenies of hominids (Figure 3.10b). Nevertheless, the arrangement of the other hominids STin2 VNTRs in this tree does not agree with previous phylogenies of hominids (Figure 3.10b). For example, it is commonly found that *H. sapiens* and the two species included in the genus *Pan* (*P. troglodytes and P. paniscus*) cluster together in phylogenetic trees. This is interpreted as these three species shared a most recent ancestor (Salem et al., 2003; Page and Goodman 2001). The species contained within the genus *Gorilla* (*G. gorilla* and *G. beringei*) often occupy the branch basal to genera *Homo* and *Pan* (Figure 3.10b). However in the present phylogeny (Figure 3.10a), the TFBS found in VNTRs of *Gorilla* sp. and *P. troglodytes* created a different branching pattern, where the VNTRs of the latter species clustered (joined by 6.33 evolutionary steps, involving changes in BS(s) for USF, AP-1, REB-1, Rev-erba and ARP-1). Nevertheless, the *P. troglodytes* and *Gorilla* sp. VNTRs exhibit a great deal of difference from each other, which indicates that their VNTRs have varied greatly since they last shared a common ancestor, as reflected by the branch length (BL) values of the branches they occupy. Indeed, the BL of the *P. troglodytes* VNTR branch is separated from the VNTR of *Gorilla* sp. by 10.66 evolutionary steps. This is produced by 6.33 evolutionary steps (involving 7 TFBS for: NF-1, Cre-1, E-1, AP2-alpha, RXR alpha, Egr-1 and CACC bind) occurring in the *P. troglodytes* STin2

VNTR branch and 4.33 evolutionary steps (involving 5 TFBS for: Cre-1, CPC-1, CRE-BP1, AP2-alpha and COUP) occurring in the *Gorilla* sp. VNTR branch.

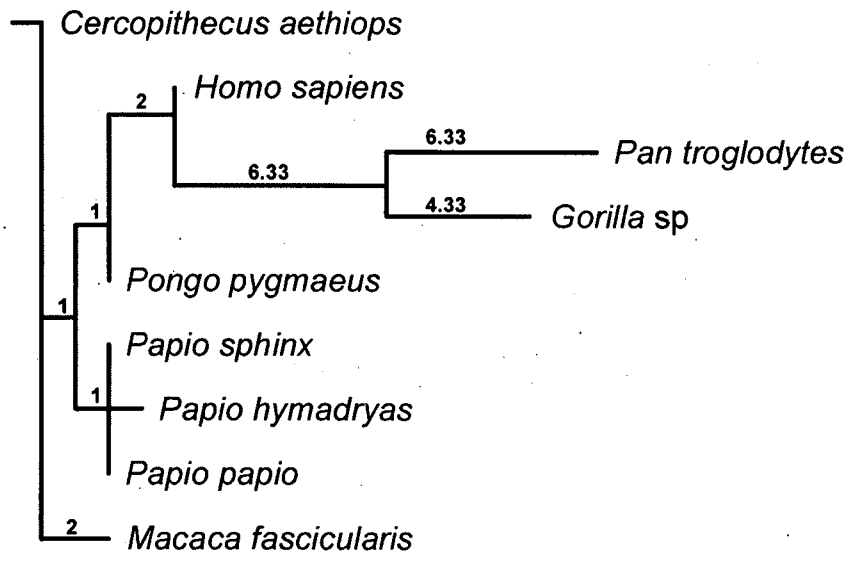
Conversely to the expected, the STin2 VNTRs of *H. sapiens* occupied a branch basal to the *P. troglodytes* and *Gorilla* sp. VNTRs branches. As the findings arising from the present phylogenetic analysis would indicate that, the VNTRs of *P. troglodytes* and *Gorilla* sp. share greater similarity in TFBS with each other than either share with the STin2 VNTRs of *H. sapiens*; the resulting tree is in conflict with the most accepted hominid phylogeny (Salem et al., 2003; Page and Goodman 2001).

I propose an explanation to explain this conflict. This concerns the nature of the evolutionary forces that occur at VNTR loci in comparison to those occurring in DNA sequences evolving by genetic drift (e.g. exons, promoters of highly conserved enhancers). The STin2 VNTR sequences are proposed to act as responsive elements to environmental stress (Klenova et al., 2004; Hariri et al., 2002a), thus it is likely that the evolution of these sequences would be more plastic than the evolution of sequences evolving by genetic drift. These latter sequences are for example those whose integrity is vital for the correct activation of a gene (e.g. a promoter or an ECR) or those which sequence determines protein structure and function (e.g. an exon). Similarly, intronic sequences that do not contain regulatory elements are also expected to change by random accumulation of mutations; however, these changes accumulate at a higher rate than observed in exons, promoters or ECRs. All these DNA sequences are selected for phylogenetic reconstructions as the changes they have accumulated are expected to portray the relationships of species, which have developed across long periods.

Given the function of the protein encoded by the SLC6A4 gene, the expression of this gene is suspected to respond to changes in the environmental conditions and stress. Therefore, if the STin2 VNTR sequences were responding to rapid environmental changes, as presumed faced by the primitive human populations during evolution, then I expected that the patterns of evolution of the STin2 VNTR reflect evolutionary changes occurring in gene expression, which are plastic in nature and do not to reflect the species ancestry. As this is the case, the variation of the patterns of evolution of the STin2 VNTRs from other type of sequences evolving by genetic drift is indicative that potential changes in the *cis* regulation of SLC6A4 indeed may have occurred during the evolution of hominids and importantly during the evolution of *Homo*.

The evolutionary reconstructions of the STin2 VNTR sequences would not necessarily contradict the accepted relationships between *H. sapiens* and *P. troglodytes*, but instead suggest that after all hominids studied last shared a common ancestor, the STin2 VNTRs of *P. troglodytes* and *Gorilla* sp. undergone the greatest divergence from the hypothetical ape ancestor. Possibly the life history similarities presented by the latter two species have resulted in common shared traits found in their STin2 VNTRs. The tree also suggest that the divergence undergone by the *H. sapiens* STin2 VNTR has resulted in VNTRs which are somehow similar to the found in the *P. pygmaeus*. The similarity between the TFBS found in these species STin2 VNTR sequences not necessarily reflects similarities in their serotonin related behaviour or similarities in the regulation of the expression of their SLC6A4 genes. This is because long time has passed since these species shared a common ancestor, therefore, it is possible that other compensating regulatory domains have evolved in the SLC6A4 genes of *H. sapiens* and *P. pygmaeus*.

a



b

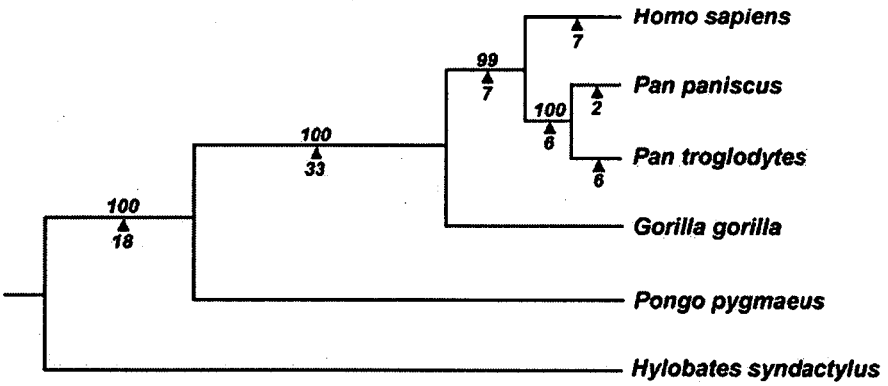


Figure 3.10 Cladograms based on the types of TFBS found in the SCL4A6 STin2 VNTR of primates and based on Alu sequences (a) In the STin2 VNTR cladogram, the African apes VNTRs (*Gorilla* sp. and *P. troglodytes*) form a cluster independent from the *H. sapiens* VNTR. The short branch length (BL, calculated using a bootstrap analysis) value that separate *H. sapiens*, *P pygmaeus* and the old world primate sequences (maximum of 3 evolutionary steps) indicates similarity amongst these sequences. **(b)** The cladogram shows accepted relationships amongst the hominids (from Salem et al., 2003). The cladogram was based on Alu elements and reconstructed by using Dollo parsimony analysis. Primate relationships were derived from analysis of 133 Alu loci by using maximum parsimony criteria. The BL values are located on top of branches in the STin2 VNTR cladogram, whereas number of insertions observed along each branch of the tree is indicated, and bootstrap support values are placed above each node.

3.3.1.7 Diversity and evolution of the SLC6A4 5' promoter VNTR

To investigate the diversity and evolution of the repeat composition of the promoter VNTR in the SLC6A4 gene of primates I aligned the repeat units of all available promoter VNTR sequences (including those amplified during this study and those found in the literature). I extended the current knowledge on the promoter 5' VNTR diversity by amplifying this region from *P. troglodytes*, *Gorilla* sp. and *P. pygmaeus* samples by PCR. The alignment produced is shown in Figure 3.11.

The alignment of the repeat units forming the promoter VNTR demonstrates that blocks of repeats have been conserved throughout evolution of hominids (Figure 3.11). For example, in the 5' end of the VNTR, 4 repeats were present in the VNTRs of *Gorilla* sp., *P. troglodytes* and *H. sapiens* ("r", "i", "s" and "t"). In the centre and 3' end of the VNTR the repeats "uvw1", "y" and "c" repeats, were found in two or more species (circled by red rectangles Figure 3.11). The intra-specific variation of copy number found in the *H. sapiens* and *Gorilla* sp. VNTRs was mainly caused by the difference in numbers of "b", "e", "d" and "k" repeats. Thus, the variation in the number of "b" and "e" repeats in *Gorilla* sp. may correlate with diversification of transcriptional activities of the promoter VNTR in this species. Interestingly, it has been shown that in modern humans, the two most commonly found VNTR copy number variants (with 14 and 16 repeats) correlated to *in vivo* and *in vitro* differential gene expression and differ in the number of "b" and "e" repeats. However, the presence of several SNPs identified in the VNTR have been shown to abrogate this transcriptional difference *in vitro* (Sakai et al., 2002).

The sequences of VNTR of *P. pygmaeus* differed greatly from the sequences of the other hominids, which created difficulties for the alignment (Figure 3.11). For example, at least 50% of repeats found in the *P. pygmaeus* VNTRs were species

specific. There was also intra-specific variation in this species. For example, whereas the 20 and 18 repeat VNTRs only differed in two repeats (repeats "b" and "e"); the VNTR with 22 repeats only shared 3 repeats ("b", "d" and "e") with other *P. pygmaeus* variants and with VNTRs of other hominids (Figure 3.11). The great variation between the repeats present in *P. pygmaeus* and other hominids 5' promoter VNTRs suggests that either the *P. pygmaeus* VNTRs have undergone recent changes or that the VNTR of *P. troglodytes*, *Gorilla* sp. and *H. sapiens* have undergone major changes after they last shared an ancestor with *P. pygmaeus*. However, more individuals of these hominid species should be analysed as VNTR variation could affect the results presented in this section. To answer some of these questions it is necessary to explore further back in evolutionary time in the cercopithecids and hylobatids (gibbons or lesser apes) 5' promoter VNTRs.

The repeat alignment also shed light on the origin of the repeat composition of the modern human promoter VNTRs. The *H. sapiens* VNTRs differed in fewer repeats with the *P. troglodytes* (3) than with the *Gorilla* sp.VNTRs (4 repeats in each copy number variant, Figure 3.11). This similarity is expected, considering that *H. sapiens* and *P. troglodytes* shared most recent common ancestor ancestry than either species does with *Gorilla* sp. This finding indicates that the changes that have accumulated in the 5' promoter VNTR repeats of hominid promoter correlate with the evolutionary distance that exists amongst hominids, unlike the observed for the STin2 VNTRs.

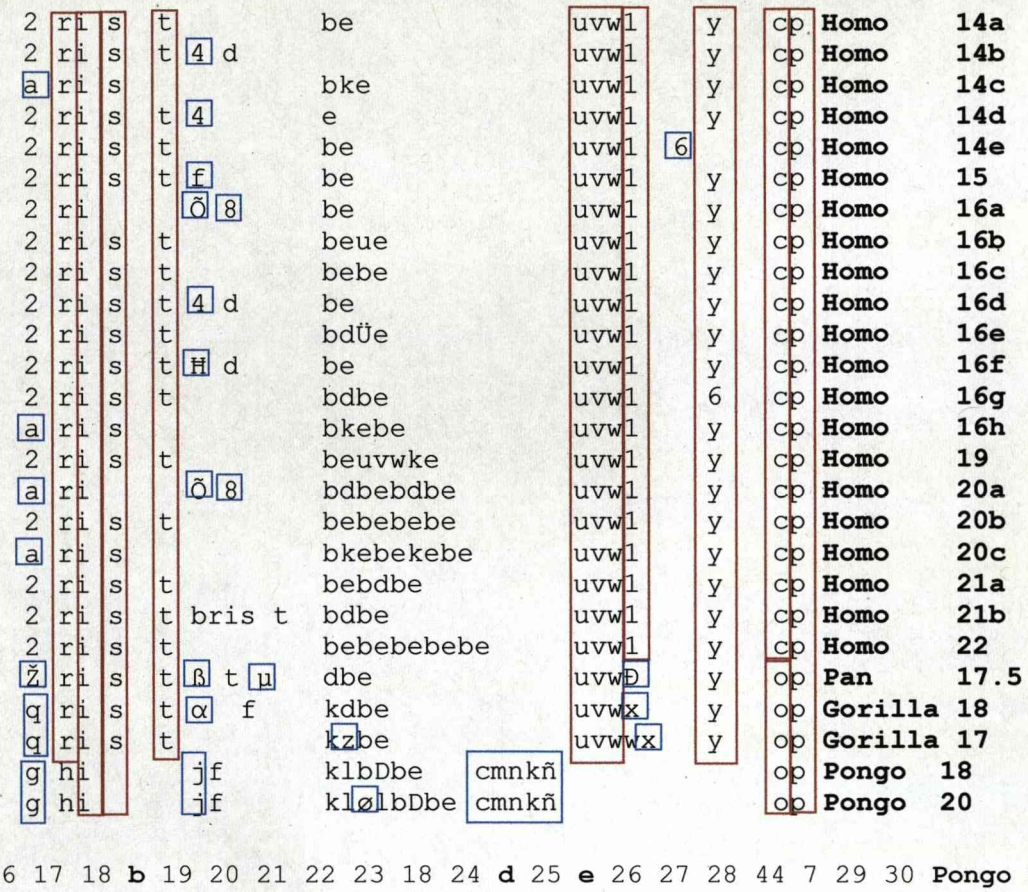


Figure 3.11 Alignment of the modern human and African ape promoter VNTR repeats. Variation in repeat conformation across species is present all along the VNTR. Intraspecific variation is predominantly present in the middle and in the 3' end of the VNTR. Between *H. sapiens* and *P. troglodytes* VNTRs, new repeats occur in the middle of the VNTR. Repeats present in more than one species are marked by red rectangles and repeats present in only one species are encircled by blue rectangles.

3.3.1.8 Old world monkeys and lesser apes promoter VNTR

In this section, I gathered the published sequences of the 5' promoter VNTR of the SLC6A4 gene of cercopithecids *M. thibetana* (n=1), *M. sylvanus* (n=1) and *M. mulatta* (n=1) and hylobatids (*Hylobates muelleri,* n=2) to generate an alignment of the VNTR repeat units (Figure 3.12). This alignment suggested these primates promoter VNTRs exhibit great repeat unit diversity. Nevertheless, specific repeats were identified all 4 species (in red font, Figure 3.12), in more than 1 species (highighted in blue font) and some other had been previously identified in the hominids 5' promoter VNTRs (marked by solid black lined rectangles). The repeats shared between cercopithecids, *Hylobates* and hominids (e.g. "k", "b", "c", "d", "e", "f" and "o") confirm that the VNTRs of these three primate groups are homologous. Moreover, due to their common features, it is likely that these repeats were present in the 5' promoter VNTRs of the ancestors to all three-primate families, around 25 mya. However, the great diversity in repeat units found in hominids and cercopithecids 5' promoter VNTR and the small number of individuals analysed impede the appropriate comparison of these VNTRs repeats.

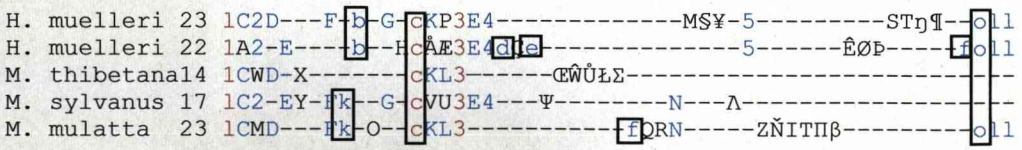


Figure 3.12 Alignment of the repeat units found in the SLC6A4 gene 5' promoter VNTRs of *Hylobates muelleri* and *Macaca* species. The number next to the species corresponds to the number of repeats forming the VNTR. Repeats circled by the solid black lines have been also seen in hominids 5' promoter VNTR. Repeats in red indicate conservation across all sequences aligned. Repeats in blue represent repeats found in more than 1 sequence and those in black font are repeats found only in 1 sequence. Spacer lines introduce to optimize the alignment.

3.3.1.9 Evolution of the SLC6A4 promoter VNTR

I analysed the promoter VNTR sequences for TFBS using the software Alibaba 2.1 as described in previous sections. I then constructed a matrix of characters (based on the TFBS found in the sequences, appendix 4.2) that was used to calculate the phylogeny. The phylogenetic analysis was based on the principle of parsimony, using the PARS program (from the PHYLIP package). The phylogenetic analysis produced the one most parsimonious tree shown in Figure 3.13a.

In the calculated tree (Figure 3.13a), the VNTRs of the cercopithecid (*M. mulatta* and *M. sylvanus*) are located on branches that have remained unresolved, i.e. their VNTRs present conflictive information. In this tree there is an internal branch formed entirely by hominids 5' promoter VNTRs (branch formed by 5 evolutionary steps). In this branch, the sequences of the *P. pygmaeus* VNTRs occupy the basal branch; and this is congruent with its positions in previous phylogenetic trees (e.g. Salem et al., 2003, Figure 3.13b). The African apes and modern human 5' promoter VNTRs share similar TFBS that clustered them in one branch. This branch is formed by 1.5 evolutionary steps. The shared traits (or TFBS) indicate that the VNTRs of these three species have not diverged greatly from the hypothetical common ancestor, unlike the observed in the tree of the STin2 VNTRs (Figure 3.10a). The arrangement of the hominids VNTRs is different from trees based on DNA sequences evolving solely by genetic drift (Figure 3.13b). For example, the 5' promoter VNTR sequence of *P. troglodytes* is basal to those of *Gorilla* sp. and to *H. sapiens.* The distance that exists between *P. troglodytes* and *H. sapiens* indicates a differentiation of their sequences with potential functional consequences. To test if the differences in TFBS found in the promoter VNTRs of *P. troglodytes* and *H. sapiens* correlate with distinct

transcriptional properties, some promoter VNTRs of the SLC6A4 gene were tested *in vitro* in chapter 4.

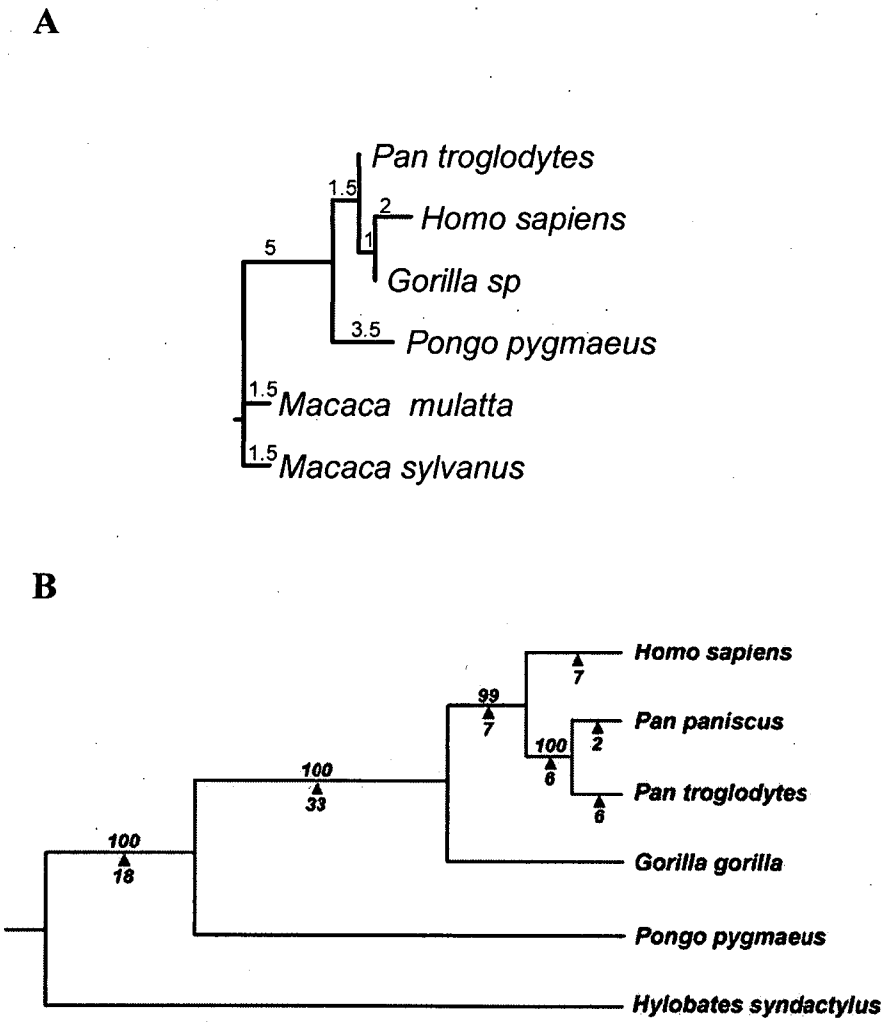


Figure 3.13 Cladograms based on transcription factor composition of the SCL4A6 promoter VNTR of primates and in Alu sequences (Salem et al., 2003). **(a)** In the promoter VNTR cladogram, the VNTRs of *H. sapiens*, *Gorilla* sp. and *P. troglodytes* form a cluster. The short branch length (BL obtained after bootstrapping calculation) value that separates them (maximum of 4.5 evolutionary steps) indicates possible recent ancestry and potential functional similarity amongst them. The BL values are located on top of branches in the promoter VNTR cladogram **(b)** Cladogram of hominids based on Alu elements and reconstructed by using Dollo parsimony analysis. Primate relationships were derived from analysis of 133 Alu loci. The number of insertions observed along each branch of the tree is indicated, and bootstrap support values are placed above each node.

3.3.2 The D4 ex3 VNTR in the DRD4 gene

The differential expression of the DRD4 gene has been associated with differences in cognitive function and personality amongst primates (modern humans, vervet monkeys) and other mammalian species e.g. horses and dogs (Bailey et al., 2007; Inoue-Murayama et al., 2002b; Larsen et al., 2005; Mogensen et al., 2006; Momozawa et al., 2005). These studies have identified correlations between specific copy number variants of a VNTR located in the third exon of the DRD4 gene (D4ex3 VNTR), behavioural and cognitive traits. However, some epidemiological studies on modern human populations have failed to replicate these correlations (Congdon et al., 2007; Lakatos et al., 2002; McCracken et al., 2000). It is possible that the high degree of polymorphism within this VNTR primary sequence could affect the functional profile of the different variants, and consequently, could affect the results of association studies. Nevertheless, no studies have explored the effect of this VNTR sequence variation to the correlation with behaviour. The D4ex3 VNTR in the DRD4 gene is present in all primates studied to date (Livak et al., 1995; Inoue-Murayama et al., 1998). The study of this VNTR in non-human primates offers opportunity to understand how this element has evolved in humans and the correlations between the evolution of the D4ex3 VNTR and the evolution of dopamine related behaviour in primates, particularly in modern humans.

Two previous studies have explored the evolution of this VNTR (Inoue-Murayama et al., 1998; Livak et al., 1995); however, these did not take into account the variability of the VNTR sequences. Therefore, in this section, I expanded the knowledge on the diversity of the primary sequence of the D4ex3 VNTR in primates and investigated the evolution of the sequence during primate evolution.

3.3.2.1 The origin of the D4ex3 VNTR in the class Mammalia

I generated an alignment of sequences (of *O. anatinus*, *G. gallus*, *X. tropicalis*, *G. aculatus*, *M. musculus*, *C. familiaris*, *M. mulatta* and *H. sapiens*) that occupy a homologous position in the DRD4 gene of vertebrates obtained from the UCSC genome browser. This alignment presented in Figure 3.14 confirms that a VNTR is present in the third exon of the DRD4 gene of *C. familiaris* (highlighted in light and dark grey) and primates (*M. mulatta* and *H. sapiens*, with repeats highlighted in yellow and brown) but is absent (48bp has not multiplied) in the DRD4 genes of non-mammal vertebrates.

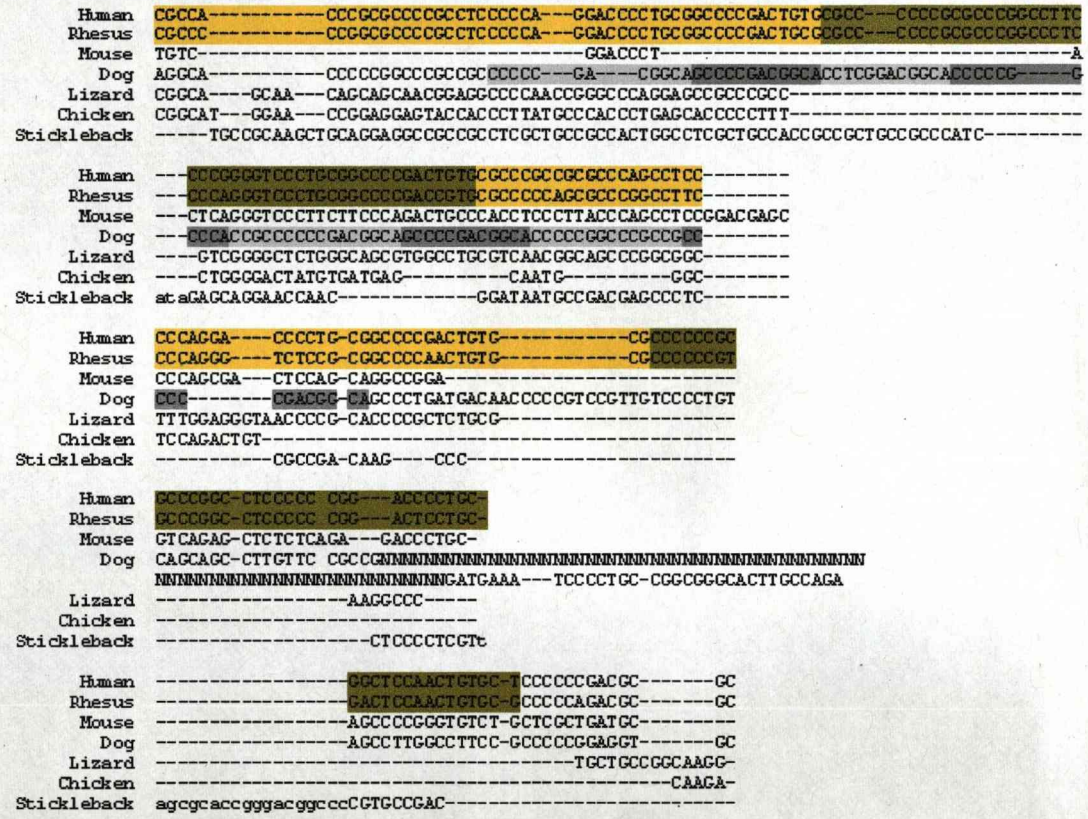


Figure 3.14 Alignment of the DRD4 exon 3 sequences of vertebrates homologous to the primate D4ex3 VNTR. Highlighted in yellow and light brown are primate D4ex3 sequences (alternating colors indicate different repeat units). Highlighted in grey are the D4ex3 repeats found in some canids (alternating grey shades indicate different repeat units)

Several studies report the presence of a VNTR in the third exon of the DRD4 gene of different mammalian groups, other than primates and carnivores mentioned above, such as ungulates and cetaceans (e.g. Livak et al., 1995; Larsen et al., 2005; Morgensen et al., 2006), however, the origin of this element in the mammalian genomes has not been investigated. I therefore searched in the NCBI database for entries of the DRD4 of different mammalian species and compared the length and sequence of their repeat units (Figure 3.15a).

In summary, the comparison of the VNTRs found in these three mammal groups demonstrate that each primate has a different VNTR consensus sequence and of different length (Figure 15a). For example, in primates the VNTR is composed by units formed by 48 bp. In some cetaceans (sperm whales, dolphins), artiodactyls (bull) and perissodactyls (horses, zebras) this VNTR is formed by repeats composed of 18 bp (consensus shown in Figure 3.15a). Finally, the DRD4 VNTR found in carnivores presents similar motifs as the VNTR in primates and ungulates (Figure 3.15.a); however, these VNTRs are formed by repeats composed by 12 bp. In the third exon of the DRD4 gene of other carnivore species such as the domestic cat (*Felis domesticus*), and in the Asiatic black bear (*Selenarctos thibetanus*) which are more distantly related to dogs and racoons, only one of the repeats has been reported (Inoue-Murayama et al., 2002b). In *F. domesticus* and *S. thibetanus* the flanking sequences of this one repeat unit found exhibited similarity with the VNTRs, which suggest that the VNTR was lost in these species after they last shared an ancestor with carnivores such as the domestic dog and racoon, around 45 mya.

The preliminary results suggest that although there is homology in the primary sequence of the VNTRs found in the third exon of the DRD4 gene of carnivores, ungulates and primates, the differences found indicate that these VNTR has evolved

independently during evolution (Figure 3.15b). The evolution of a VNTR in the third exon of the DRD4 gene of different mammals may indicate that this exonic locus is a hotspot for recombinant events, and furthermore, may be correlated with the diversification of personality traits in the class Mammalia.

a

Equus caballus
DRD4 exon 3 VNTR

```
CACCCCTGCCTCCGACGC
CACCCCGCCCCCGACGC
CACCGCACCCCCTGACGC
CACTCCGCCCCCTGACGC
CATCCTGTACCCCGACGC
CACTGCACCCCCTGACGC
CACGCTGCCCCGCGACGC
CATCCTGTCCCCGGATGC
CATCCCGTCCCCCGATGC
CGTCGCACCCCTCCACGC
CATCCCGGCCGAGCCCCC
```

Canis familiaris DRD4 exon 3 VNTR

```
CCCCCGACGGCA
GCCGCGACGGCA
CCTCGGACGGCA
-CCCCCGGCCCA
-CCGCCCCCCGA
CGGCAGCCCCGA
```

Pongo pygmaeus DRD4 exon 3 VNTR

```
CCCCGCCCCCGCCTCCCCAGGACCCCTGCGGCCCTGACTGTGCGCC
CCCCGCCCCCGCCTTCCCCAGGGTCCCTGTAGCCCCAACTGCGCGTC
CCCAGGGCCCGGCCTTCCCGGGGGTCCCTGCGGCCCCGACTGTACGCC
CCCCGCGCCCCGGCCTCCCCCCGGACCCCTGCGGCAACTGTGCGCCGCC
```

b

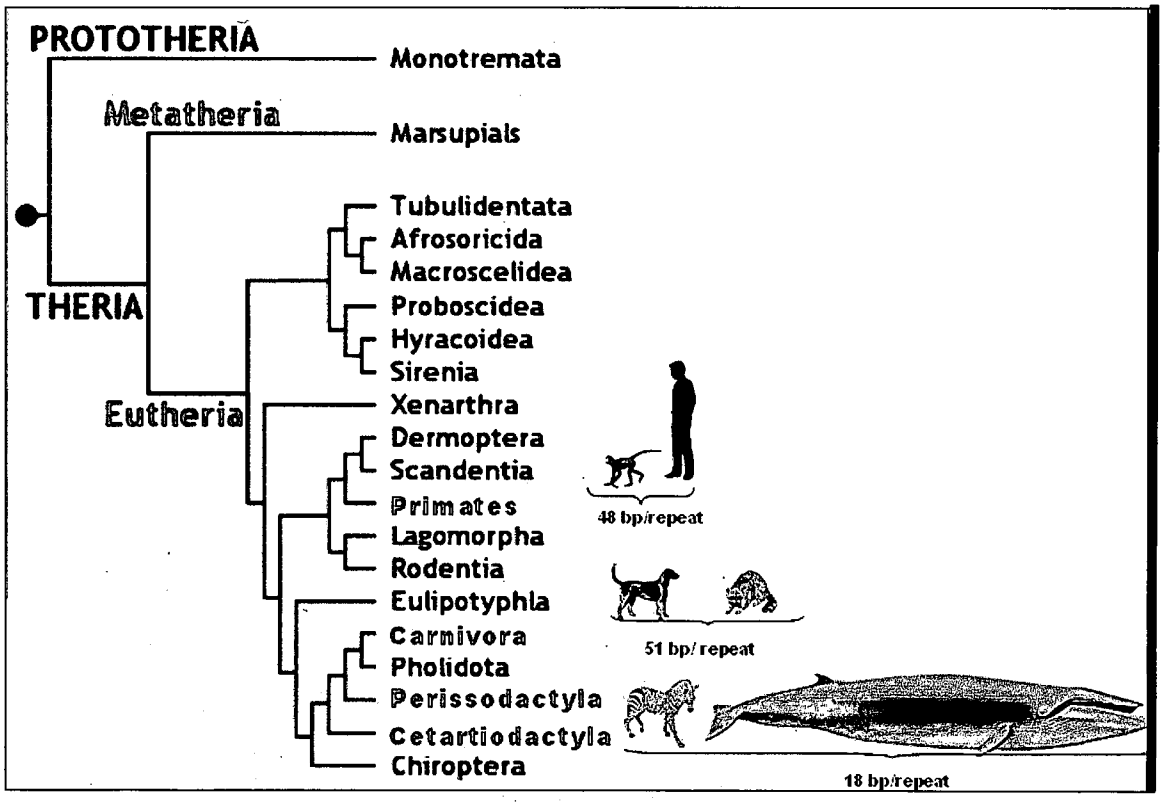


Figure 3.15 The D4ex3 VNTR of the DRD4 gene of mammals (a) The VNTR in ungulates (*E. caballus*). and canids (*C. familiaris*) show sequence homology (examples are highlighted in grey) to the primate D4ex 3 VNTR **(b)** Groups of mammals that present a VNTR in the third exon of the DRD4 gene. Diversity of the D4ex3 VNTR in mammals.

3.3.2.2 Diversity and evolution of the D4ex3 VNTR in primates

It is recognised that the cognitive abilities of the primates are greater in modern primate groups such as the apes and old world monkeys than the found in prosimians or new world monkeys. The D4ex3 VNTR found in the DRD4 gene has been associated to cognitive and behaviour of modern humans (Demiralp et al., 2007; Ding et al., 2002; Gornick et al., 2007). This has been attributed to the effect the polymorphism has on the length and function of the 3rd cytoplasmatic loop of the receptor (Demiralp et al., 2007) and its also proposed function as a *cis* regulator of gene expression (Schoots and Van Tol, 2003).

To analyse correlation between cognitive evolution in primates and the evolution of the D4ex3 VNTR sequences, I studied the diversity of this VNTR in primates and in particular changes in the D4ex3 VNTR sequences, which have occurred during the evolution of the primate lineage.

I first expanded on the knowledge of the diversity of the D4ex3 VNTR of primates I amplified by PCR (Figure 3.16). These PCR amplifications demonstrated that this VNTR region is present in new world monkey species not studied before (*Ateles paniscus* "black spider monkey", n=1; *Lagothrix lagotricha* "woolly monkey" n=1 and from *Saguinus fuscicollis* "common saddle back tamarin" n=1). I also amplified the VNTR region from species previously studied (*P. troglodytes*, *Macaca mulatta*, *Lemur catta*, *Tarsius bancanus*) to increase the number of samples analysed. The amplified sequences are shown in appendix 5. In brief, the PCR amplification revealed the presence of novel copy number variants of the VNTR in the prosimian *Lemur catta* (with 3 and 4 repeat units per VNTR) and in the tarsier *Tarsius bancanus* (with 1 repeat unit). In the new world monkeys, the *Lagothrix lagotricha* individual was heterozygous for the VNTR region, carrying 3 and 4 repeat units per VNTR. The

Ateles paniscus and *Saguinus fuscicollis* individuals were both homozygous, and carried VNTRs with 4 and 5 repeat units respectively.

All of the known sequences of the D4ex3 VNTR of primates found to date (including those obtained by PCR amplification in this study and those found in the literature and in the NCBI database) are summarised in Table 3.2. Briefly, it was observed that the D4ex3 VNTR of the Hominidae (modern humans and great apes) presented greater variation of copy number (from 2 to 11 repeat units per VNTR) than any other primate group. Interestingly, the greatest variability is exhibited by *H. sapiens*, which carries the VNTR with the lowest (2) and highest number (11) of repeat units. However, the number of individuals studied from *H. sapiens* is greater than the number studied from any other primate; therefore, to validate this finding more individuals from other species must be investigated.

In cercopithecids, 3 species studied had both 4 or 5 repeat units per VNTR (*Macaca mulatta*, *M. fascicularis* and *M. tonkeana*); 3 species presented only 5 repeats (*M. nemestrina*, *M. thibetana* and *M. arctoides*) and 1 presented only 4 repeats per VNTR (*M. sylvanus*). In the least explored group, the new world monkeys, the copy number varied greatly across species, ranging from 3 (In *Ateles paniscus*) to 9 repeat units (in *Saimiri boliviensis*) per VNTR. In *Tarsius bancanus*, a representative of tarsiers, the VNTRs presented from 1 to 9 repeats per VNTR (Table 3.2).

In the most primitive primate group, the prosimians, the VNTR copy number did not vary as greatly as observed in the other groups, and the number of repeat units per VNTR was lower than in other primates. For example, *Nycticebus caucang* carried VNTRs with 1 or 2 repeat units and *Lemur catta* carried VNTRs with 1, 2 and 3 repeat units.

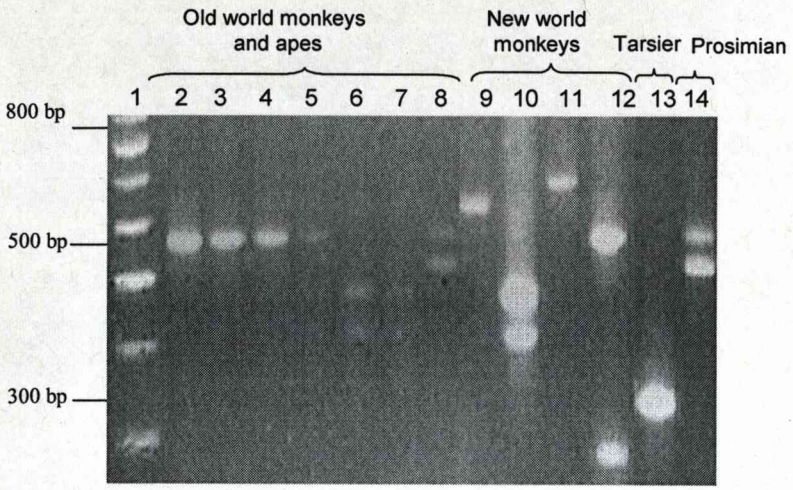


Figure 3.16 PCR amplification of the D4ex3 VNTR of primates. Lane1: 100 bp marker, lanes 2-5 *P. troglodytes*, lane 6: *M. sphinx*, lane 7 *C. aethiops*, lane 8: *M. mulatta*; lane 9 *Lagothrix cana*, Lane 10 *Ateles belzebuth*, lane 11: *Cebus apella*, lane 12: *Saguinus fuscicollis* lane 13: *Tarsius bancanus,* lane 15 *Lemur catta.* Heterozygous VNTR were observed in all primate groups. However all old world monkeys and apes VNTRs analysed presented more than 3 copies repeats whereas the tarsier only presented 1 repeat unit.

				Sample size	Number of repeat units per VNTR	Total number of TFBS per VNTR	Different types of TFBS per VNTR
INSECTIVORA			Tupaiidae	3	1	2	1
PRIMATES	Strepsirhines		Lemuridae	4	2-4	12-30	6-9
			Galagonidae	2	1-2	5-17	2-7
	Haplorhines	Tarsiers	Tarsiidae	2	1-9	37	13
		Anthropoids	Callitrichidae	1	5	38	13
			Cebidae	4	3-9	26-99	7-9
			Cercopithecidae	275*	4-5	26-48	9-13
			Hylobatidae	16	4-8	33-64	13-14
			Hominidae	22	2-11	25-84	7-13

Table 3.2 Sequence variation of the VNTR in the third exon of the DRD4 gene of primates. Primates are divided in 2 main groups: Haplorhines, including tarsiers and anthropoids and Strepsirhines, including all prosimians. In this table, I also included a shrew (Order Scandentia), *Tupaia glis,* which represents closest living relative to all primates. * indicates data obtained from Wendland et al., 2006b.

3.3.2.2.1 Correlation between the length of the D4ex3 VNTR and the evolution of primates

To investigate whether the observation where prosimians have shorter VNTRs than other primates (Table 3.2) is significant. I calculated the mean, standard deviation and standard error of the number of repeat units found in the VNTRs of each primate group (apes, old world and new world monkeys and prosimians) and plotted these data in Figure 3.17. In this graph the tarsiers were not included because the sample size was n=2. In brief, the graph shows that the number of repeat units per VNTR is different between anthropoid primates and prosimians. Albeit the variation found in each primate group demonstrated (as reflected by the standard error bars), the prosimians VNTR have significantly lower number of repeat units than any other primate group analysed (Students' *T*-test shows anthropoids vs. new world monkeys $p<0.005$; vs. old world monkeys $p=0.05$ and vs. hominids $p<0.005$). The graph further indicates that amongst anthropoids the mean number of repeat units is not significantly different (Figure 3.17). This result suggest that the exon 3 VNTR of the DRD4 gene in primates became longer with the evolution of anthropoids, approximately 40 mya (Gunnell and Miller, 2001). The difference in the length of the prosimians and anthropoids D4ex3 VNTRs suggests potential transcriptional differences and functional differences of their receptor. However, the sequence variation observed in the VNTR could affect these transcriptional properties, thus *in vitro* studies are required to further this hypothesis. To further investigate the functional differences generated by the sequence variation, in the following section I compared the diversity of putative TFBS found in the sequence of the D4ex3 VNTR of prosimians and anthropoids, as a proxy for potential functional differences between the VNTRs of these primate groups.

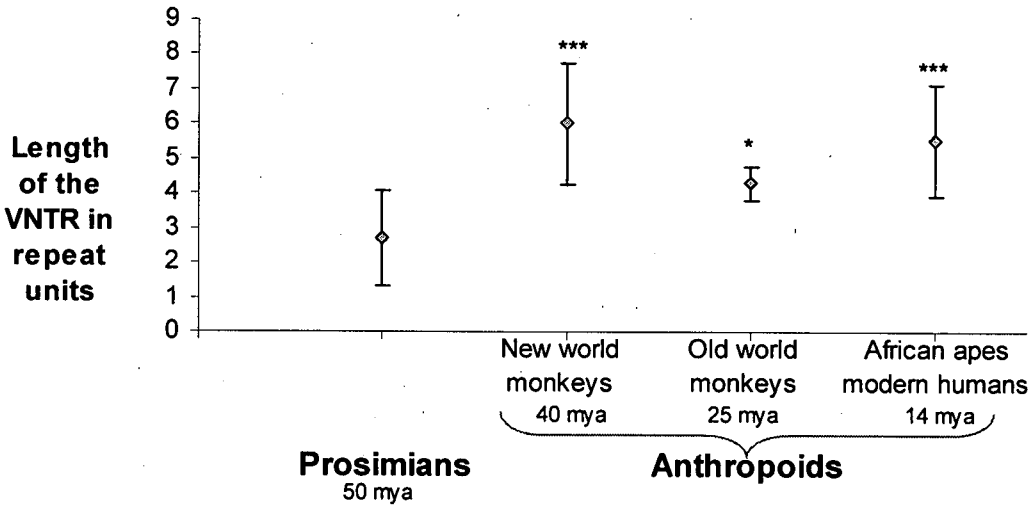


Figure 3.17 Number of repeat units forming the VNTR D4ex3 of the DRD4 gene of primates is higher in anthropoids than in prosimians. The average D4ex3 VNTR length (expressed in number of repeat units) of each primate group is significantly lower in prosimians than in new world monkeys (Student's *T*-test, ***= $p<0.005$), than in old world monkeys (*=$p\leq0.05$) and than in great apes (***=$p\leq0.005$). The error bars reflect the variability of the number of repeat units found forming the VNTRs of each primate group.

3.3.2.2.2 The number of TFBS in the D4ex3 VNTR increases with the evolution of primates.

In the previous section, it was shown that the average length of the VNTRs of the exon 3 VNTR of the DRD4 gene of prosimians was shorter than that of anthropoids. To investigate further how length effect could modify transcriptional properties of primates VNTRs, I conducted a comparison of the diversity of the TFBS found within the VNTRs sequences found in each primate group. The TFBS were identified by the program Alibaba 2.1, based on the TRANSFAC 4.0 database. The list of TFBS found in each VNTR sequence is detailed in appendix 6.1

In brief, Figure 3.18 shows that the diversity of TFBS found in the prosimians and anthropoids is different. In the prosimians, the mean of different types of TFBS is 6.2, whereas in new world monkeys, old world monkeys, African apes and modern humans this mean is higher than in prosimians (≈10.4 TFBS). Statistical analysis (Student's *T*-test) shows that the difference between prosimians and each other primate group studied is very highly significant (p=0.005 for the 3 comparisons). The results also suggest that the diversity of TFBS is not different amongst all anthropoids. However, the intra-group variation of the diversity of TFBS is smaller (as reflected by the small error bars) than the variation of the number of repeats found in the VNTR in the previous section. This finding indicates that the diversity of TFBS found within their sequences is highly conserved within each primate group. This may be caused by the increase purifying selection pressures that typically apply on the exonic sequences to preserve the amino acid sequences of proteins (Santini et al., 2003). Nevertheless, the results suggest that the major differentiation between the TFBS of the D4ex3 VNTRs appeared 40 mya, with the birth of the anthropoids, and

offers further support to the hypothesis that these VNTRs may contribute to the variation of the transcriptional profile of the DRD4 gene in primates.

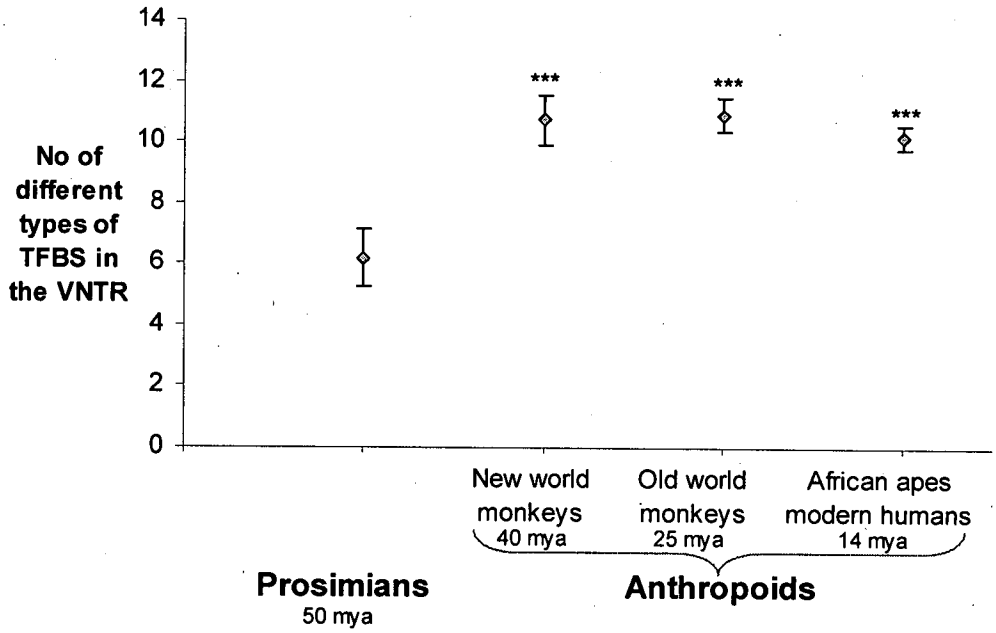


Figure 3.18 The diversity of the D4ex3 VNTR is greater in the anthropoids than in prosimians. The average number of different TFBS found prosimians D4ex3 VNTRs is significantly lower than the found in new world, old world monkeys, African apes and modern humans D4ex3 VNTRs (Student's *T*-test where *** indicate $p<0.005$). The difference in the average number of TFBS found amongst anthropoid primates D4ex3 VNTRs is not significant.

3.3.2.2.3 Number of repeat units can correlate with the diversity of TFBS found in the D4ex3 VNTRs.

The similarities between graphs describing the distribution of the length of the VNTRs and diversity of the TFBS in the different groups of primates suggest that these two variables may be positively correlated. This correlation is expected as the longer the DNA sequence, the more TFBS are likely to appear in it. To investigate this further, I plotted the average number of different type of TFBS found within each VNTR versus the length of the D4ex3 VNTR expressed in number of repeat units. The graph produced in shown in Figure 3.19.

The results in Figure 3.19 suggest that the diversity of the types of TFBS dramatically increases between 1 and 3 repeat units. When the index of correlation between these data points (from 1 to 3 repeats per VNTR) was calculated (using a polymomial correlation, equation $y = -0.9583x^2 + 7.9583x - 5.75$), this value was found to be highly significant ($R^2=1$). This finding indicates that the increase of the diversity of TFBS is strongly correlated with the increase in VNTR length between 1 to 3 repeat units. The same calculation was applied to the rest of data points; however, the correlation between the diversity of TFBS and the length of the D4ex3 VNTR was not significant ($R^2 = 0.0286$). A student's *T*-test showed that the average number of TFBS in the sequence with 1 and 2 repeat units was significantly lower than the VNTR with 3 or more repeat units (Figure 3.19). The average number of TFBS of VNTR with 3 or 4 repeats was lower than VNTRs with more repeat units, however, the presence of abundant variation within each VNTR length category (as reflected by the error bars) caused this difference to be not significant.

The present findings suggest that the multiplication of copies of the repeat units may contribute to enhance the diversity of binding sites in this VNTR,

increasing the interactions between the TF and the regulatory domain. This maybe one factor which contributes to the increase of neurotransmission complexity, proposed to occur in animals with high cognitive capacities like anthropoids (Reader and Laland, 2002). However, in the case of the D4ex3 VNTR, beyond three repeats, the diversity of TFBS seems to be a consequence of the primary sequence variation than a consequence of the VNTR length.

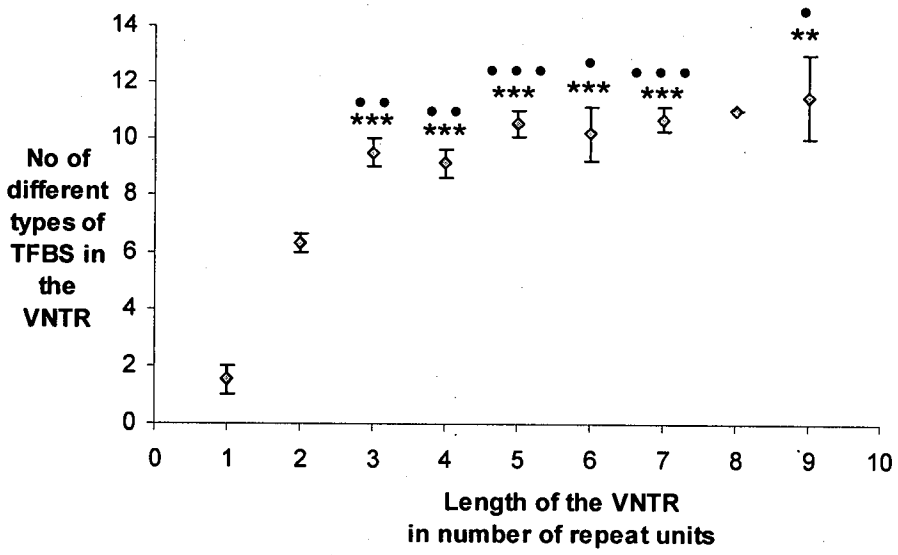


Figure 3.19. Correlation between the diversity of TFBS within the D4ex3 VNTR and its length. The number of types of TFBS in the sequence of a VNTR formed by 1 or 2 repeats were compared to the number of types of TFBS found in VNTRs with 3 or more repeats. The number of TFBS in the D4ex3 VNTR significantly increases between 1 and 3 repeats (R^2=1). Students' *T*-test demonstrated significant differences between the number of types of TFBS found in VNTRs with 3 repeats or more and the diversity of TFBS found in VNTRs with 1 or 2 repeat units (indicated by * and • respectively).

3.3.2.3 Evolution of the D4ex3 VNTR in primates

In this section, I analysed the evolution of the primate D4ex3 VNTR sequence to determine further changes occurring at the species levels, particularly to better understand events that may have predated the evolution of the modern human D4ex3 VNTR. The possible implications on the D4ex3 VNTR with the evolution of primate cognition and personality encouraged the study of the evolution of the DRD4 VNTR (Livak et al., 1995, Inoue-Murayama et al., 1998; Baily et al., 2007). However, the great variability in the VNTR sequence has not permitted the use of alignments of the entire sequence for phylogenetic reconstructions. Instead, these studies have used alignment of only the first repeat (Livak et al., 1995) or have analysed the evolution of each repeat unit individually forming the VNTRs (Inoue-Murayama et al., 1998). Moreover, the functional evolution of this VNTR has not been explored. For this reason, I conducted a phylogenetic reconstruction of the evolution the D4ex3 VNTR of primates, based on the different TFBSs identified in all known D4ex3 sequences (appendix), as a proxy for functional evolution of this D4ex3 VNTR sequence. The TFBS were coded as characters for the building of a matrix as described in sections 2.2.19.2. The matrix produced is shown in appendix 6.2. The phylogenetic analysis was based on the principle of parsimony and was executed using the parsimony program PARS of PHYLIP (phylogenetic inference package).

The phylogenetic analysis produced one most parsimonious tree (82 evolutionary steps, Figure 3.20a). The branching pattern found in the D4ex3 VNTR tree differs greatly from the known phylogenetic relationships of the primate observed in previous studies based on exonic sequences (e.g. based on exon 1 of the interstitial retinoic binding protein, Poux and Douzery, 2004 Figure 3.20b). For example, the prosimian VNTR sequences occupy the most basal branches; however, the sequence

variants of one species do not cluster together. Within the tree, all anthropoid VNTR sequences occupy one branch. The basal branch to all hominids is occupied by the VNTR of *Tarsius bancanus*, a species recognised as a representative of the most ancient group of anthropoids. Most old world monkeys and apes VNTRs are distributed in different branches; however, one of the VNTR variants of the prosimian *Lemur catta* was positioned amonsgt the cercopithecids sequences.

Inoue-Murayama (1998) demonstrated that primary sequence variation that exists amongst some primates (prosimians and tarsiers) D4ex3 VNTRs affect the amino acid sequence of the D4 receptor. Therefore, the lack of similarity between the primates D4ex3 VNTR phylogeny and others based on exonic sequences suggest that this VNTR could affect the conservation of the function of DRD4 protein across the primate order.

Within the anthropoid branch, the VNTR variants of hominids species do cluster in species-specific clusters. The separation of most anthropoid and prosimian sequences in the resulting tree suggests that there is a degree of conservation, perhaps because these sequences encode for aminoacids. Finally, the resolution of the hominid VNTR branches within this cladogram suggests that the TFBS of a putative regulatory domain can provide useful information for studying the evolution of the D4ex3 sequences. I further investigated this by reconstructing the evolution of the hominids D4ex3 VNTRs.

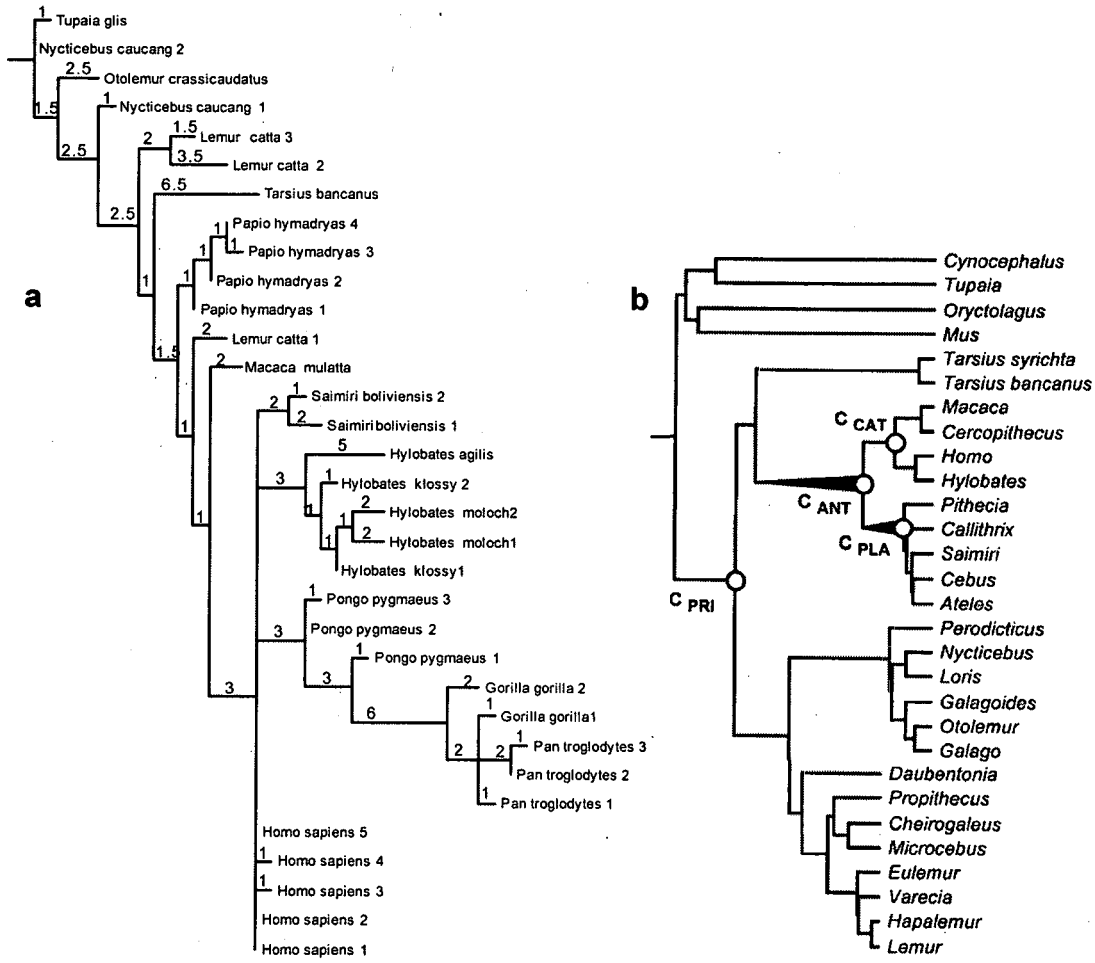


Figure 3.20 Comparison of a cladogram based on the D4ex3 VNTR of primates based and one based on the first exon of the interstitial retinoid binding protein IRBP. (a) In the D4ex3 VNTR tree, the relationships amongst the species are not similar to those found when the tree is calculated using exonic sequences, that evolved by genetic drift e.g. IRBP tree **(b)** Extracted from Poux and Douzery (2004).

3.3.2.4 Evolution of hominids D4ex3 VNTR

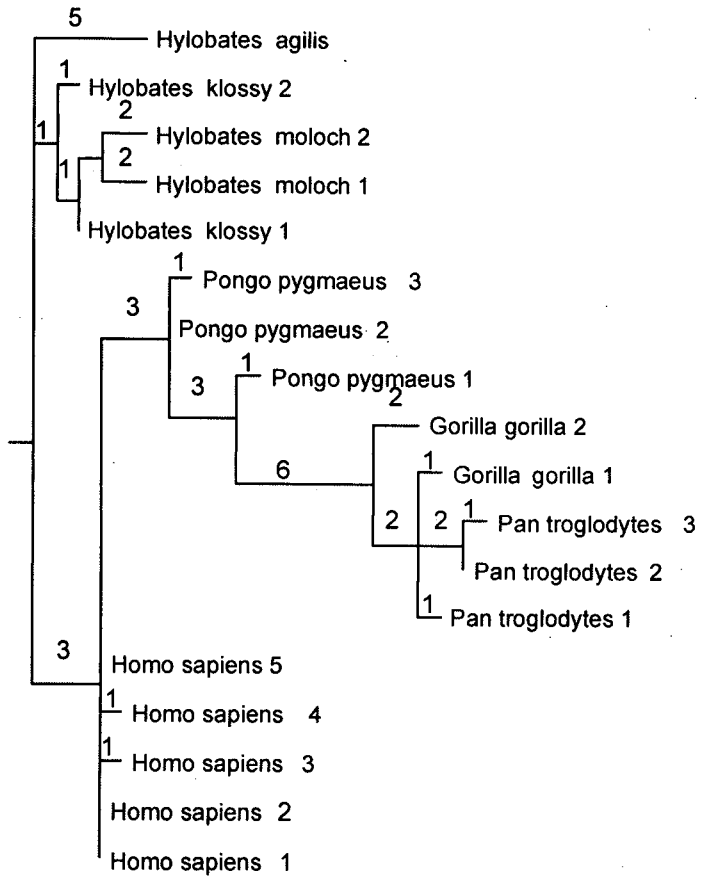
Phylogenetic analysis of the D4ex3 VNTR of hominids produced one most parsimonious tree (41 steps long, Figure 3.21a). For the construction of this tree, the sequences of the genus *Hylobates* (lesser apes) were used as outgroup. In the resulting tree, the great apes cluster in one separate branch (length of the branch or BL =3, involving changes in TFBS for YY1, Oct-1 and Adf-1). Interestingly, in this tree the BL of the branches occupied by the *H. sapiens* VNTR variants is 0 or 1 (from the hominid node), and occupy a basal position to the D4ex3 VNTR variants of *P. pygmaeus*, *Gorilla* sp. and *P. troglodytes*. The position and length of the *H. sapiens* VNTRs indicate that their sequences do not differ greatly from the sequences of the hypothetical hominid ancestor D4ex3 VNTR or from the VNTRs of *Hylobates*. This suggests that the *H. sapiens* VNTRs have differentiated little from the TFBS found in the hypothetical ancestor to all hominids. The clustering of *P. pygmaeus*, *Gorilla* sp. and *P. troglodytes* VNTR sequences is created by 3 evolutionary steps (BL=3) involving TFBS for MIG, CEB/Palp and WT1. This result suggests that the TFBS in these hominids VNTRs have differentiated further from the hominid ancestor D4ex3 VNTR.

In the clade formed by the great ape VNTRs, the *P. pygmaeus* VNTR variants were positioned basally to the *Gorilla* sp. and *P. troglodytes* VNTR branches (Figure 3.21a). The BL of the branch from where the *Gorilla* sp. and *P. troglodytes* branches stem is the longest in the tree, and is formed by 6 evolutionary steps (involving TFBS for YY1, USF, NFmμE, MIG, GATA-1 and NF-1). The branching pattern indicates that the D4ex3 VNTRs of *P. troglodytes* and *Gorilla* sp. share TFBS, which distinguish them from the *P. pygmaeus* VNTRs. However, there is intra-specific variability in the TFBS of *Gorilla* sp. and *P. troglodytes*, as reflected by the BL of

their internal branches. The topology of this hominid tree, where *H. sapiens* occupies a basal position with respect to the rest of hominids is incongruent with the accepted phylogenies of this group (e.g. Salem et al., 2003 Figure 3.21b). The most parsimonious explanation for the basal position of the human VNTRs in this tree, and the further divergence undergone by the VNTRs of all other hominids is that after the split from the hominids ancestor, the sequence of the *H. sapiens* D4ex3 VNTR variants have changed at least once more than the other hominids. This extra change would have generated the similarity between the modern human and the ancestral hominid VNTR TFBSs. Furthermore, I propose that as observed in the resulting tree, the variation in the TFBS of *H. sapiens* and the other hominids D4ex3 VNTR suggest potential functional divergence and this may have consequences on their transcriptional properties on DRD4 gene expression and the D4 receptor function.

Finally, the present results suggesting functional and evolutionary divergence of the *H. sapiens* and other great ape D4ex3 VNTRs complement the results of the phylogenetic analysis of the hominid STin2 VNTR in the SLC6A4 gene (Figure 3.10a). The consequences of the D4ex3 VNTRs sequence divergence are explored *in vitro* in chapter 4.

a



b

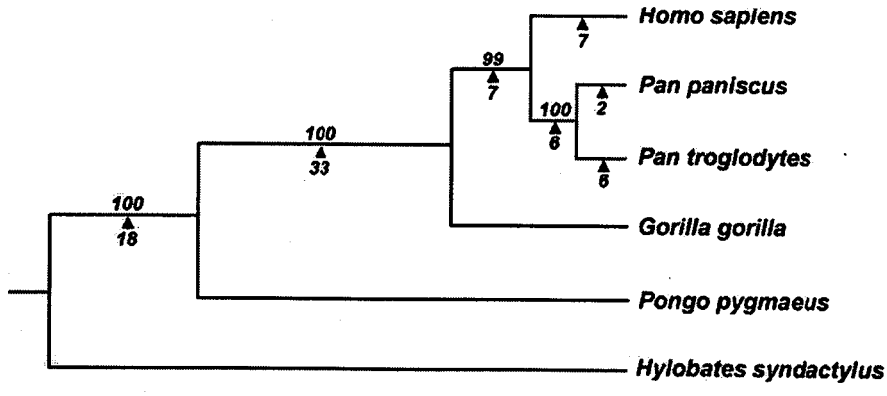


Figure 3.21 Comparison of the evolution of the D4ex3 VNTR in hominids based on their TFBS and a cladogram based on alu sequences. (a) In this cladogram the D4ex3 VNTR sequences of the great apes clustered in a branch separate from the *H. sapiens* D4ex3 sequences. The number of evolutionary steps for each node is positioned on top of each branch. **(b)** Cladogram of the hominids based on Alu sequences (Salem et al., 2003) which evolve by genetic drift. The number of insertions observed along each branch of the tree is indicated, and bootstrap support values are placed above each node.

3.4 Brief summary and discussion

This chapter aimed to investigate questions regarding the origin and diversification of VNTRs in the DRD4 and SLC6A4 genes. Sections 3.3.1.1 and 3.3.1.2 focused on investigating the origin of the 5' promoter and STin2 VNTRs of the SLC6A4 gene in the genomes of mammals and other vertebrates. The cross species comparisons demonstrated that a tandem repeat with great homology for the primate 5'promoter VNTR is present in the serotonin transporter genes of vertebrates (Figure 3.5), whilst no repetitive elements homologous to the STin2 VNTR of primates were found in the serotonin transporter gene of vertebrates (Figure 3.2). These results indicate that the region proximal to the 5' promoter of the SLC6A4 gene of vertebrates is a hotspot for recombination, perhaps reflecting that this region typically contains variable elements, which correlate with diversification of emotional behaviour and stress response in all vertebrates. The absence of this VNTR in the 5' promoter region of the serotonin transporter gene of rodents has been interpreted as that this VNTR is exclusive to primates. However, it is possible that this element is present in the promoter regions of the serotonin transporter genes of other mammalian species not investigated in this or previous studies.

The absence of a VNTR in the second intron (STin2) of the SLC6A4 gene from the genomes of vertebrates and non-primate mammals suggest that this VNTR is an innovation of the primate genomes, which increment the *cis* regulatory complexity of their SLC6A4 gene expression. The appearance of this element in the genomes of primates may correlate with the increase complexity of the behavioural responses to other individuals in their intricate social groups (Dunbar, 2003; Dunbar and Bever, 1998). Nevertheless, it is important that more mammalian species be explored to validate this hypothesis.

In silico examination of the third exon of the D4ex3 VNTR of non-mammalian vertebrates demonstrated that this VNTR is absent from their genomes (Figure 3.14). However, this VNTR has been reported in other mammal groups such as carnivores, cetaceans and ungulates (Larsen et al., 2005; Mogensen et al., 2006). Comparison of the sequence of the D4ex3 VNTRs of these mammals shows that all VNTRs in the third exon of the DRD4 gene are high in GC content and present sequence homology to the primate VNTR, however, organization of the VNTR is different in each group (Figure 3.15a). This suggests that a VNTR has appeared *de novo* in the third exon of the DRD4 gene of some species of carnivores, ungulates, primates and cetaceans. The variation created by this exonic VNTR may have an effect on the function of the D4 receptor or affect the expression of this gene, as proposed for the *H. sapiens* D4ex3 VNTRs (Schoots and Van Tol, 2003).

The second aim of this chapter was to investigate the diversity and evolution of the VNTRs of the DRD4 and SLC6A4 genes in primates. Comparison of the diversity of the number of repeat units forming the STin2 VNTR (Figure 3.6) suggested that the sequences have diversified greatly in hominids (VNTRs with 6 to 40 repeat units), but interestingly this diversity is lower in the cercopithecids which exhibit 5 repeat unit per VNTR in every species analysed. This suggests that specific selective pressures have may result in the great diversification of the hominid VNTRs. This is confirmed by the study of the evolution of the TFBS within the VNTR sequences (Figure 3.10a). This analysis showed that whereas TFBS in the *H. sapiens* and *P. pygmaeus* VNTRs were not very different from each other or from the cercopithecids VNTRs, the TFBS in the VNTRs of *P. troglodytes* and *Gorilla* sp. have diverged greatly from those found in other primate STin2 VNTRs. I propose two hypotheses to explain these findings. The first poses that after their split from

ancestral hominid, the VNTRs of the ancestor to African hominids (*Gorilla* sp., *P. trolgodytes* and *H. sapiens*) expanded and the ancestry shared by *P. troglodytes* and *Gorilla* sp.would explain the common TFBS in their sequences (as represented in Figure 3.10a). In this scenario, the STin2 VNTRs of *H. sapiens* would have reduced in length and in this process losing the TFBS shared with *P. troglodytes* and *Gorilla* sp. However, the most parsimonious explanation suggest that these three species separated from the common hominid ancestor, the STin2 VNTRs of *P. troglodytes* and *Gorilla* sp. expanded and share similar TFBS due to parallels in their habitat and behaviour e.g. Dainton and Macho, 1999.

The alignment of the 5'promoter VNTRs repeats showed that although there is variation amongst hominids VNTRs, many specific repeats were conserved across all species (Figure 3.11). However since the sample size used in this study is small, more promoter VNTR variants should be studied to validate this hypothesis. The phylogenetic analysis of the promoter VNTRs of hominids (Figure 3.13a) revealed that all hominids share many similar TFBS; and that the turnover of these TFBS is slower than seen in the STin2 VNTRs (Figure 3.10a). Unlike the STin2 VNTR, the topology of the promoter VNTR tree is similar to that of trees that are based on sequences evolving by genetic drift (e.g. Salem et al., 2003 Figure 3.10b). However, it was seen that there is a difference between the TFBS found in the *H. sapiens* and *P. troglodytes* promoter VNTRs; therefore suggesting potential functional differences amongst these species VNTRs.

The STin2 VNTR has evolved faster than the 5'promoter VNTR has in primates, which would suggest that both VNTRs are subject to different selective pressures. It is possible that the location of the 5' promoter VNTR in the SLC6A4 gene would reduce the accumulation of changes in this region, because of physical

proximity and therefore interactions with other conserved regulatory domains in the promoter region. As suggested by the presence of TRs at the 5' region of the serotonin transporter gene of vertebrates (Figure 3.5) it is also possible that this promoter VNTR may be functionally important for the regulation of the serotonin transporter gene in vertebrates, therefore, great variability at this locus would be reduced by purifying (negative) selection.

The distance and location of the STin2 VNTR from the 5' proximal promoter of the SLC6A4 gene may have allowed rapid change and turnover of TFBSs. This may indicate that the STin2 VNTR exhibits greater functional plasticity to regulate the expression of the SLC6A4 gene than the 5' promoter VNTR does. Therefore, whenever there is a change in the environmental conditions faced by a primate, the SLC6A4 gene expression could be more readily change via modulation of the STin2 VNTR. The rapid "evolvability" of the STin2 VNTR locus must not be interpreted as lack of function, as its activity has been demonstrated *in vitro* and *in vivo* (Klenova et al., 2004; Mackenzie and Quinn 1999). Furthermore, this STin2 VNTR is clinically important (e.g. Battersby et al., 1996) and previous studies have suggested that this VNTR may have contributed to the evolution and adaptability of modern humans (Gelernter et al., 1999). In addition, the differences in the rate of evolution exhibited by the two VNTRs of the SLC6A4 gene of primates suggest that these two regulatory domains may exhibit some variation in their function under some specific cellular conditions or during specific developmental stages.

In this chapter, I analysed whether there was a correlation between the length and diversity of the D4ex3 VNTR sequence and the evolution of cognition in primates (Figure 3.17). The analysis demonstrated that the diversity of TFBS is greater in the group of primates, which exhibit more advanced cognitive abilities such as

anthropoids, than in the group with less advanced abilities such as prosimians (Roth and Dicke 2005). However, although these results suggest a correlation between the diversity of the TFBS in the D4ex3 VNTR (as a proxy for *cis* regulation complexity) and primate cognitive capacities as stated earlier, this difference could have been caused by the differences of length of the VNTR. In VNTRs with less than 3 repeat units (as commonly seen in prosimians) the diversity of TFBS was significantly lower than in VNTRs with more repeats (Figure 3.19). In conclusion, I propose that the increase in the length of this VNTR potentially correlates with increase in diversity of putative TFBS, used as a proxy for complexity of the *cis* regulation of the DRD4 gene expression.

Finally, I analysed the evolution of the D4ex3 VNTR in the third exon of the DRD4 gene of hominids (Figure 3.21a). This analysis showed that this VNTR had undergone similar evolutionary pathway as seen for the STin2 VNTRs, where the sequences of *H. sapiens* D4ex3 VNTRs present markedly different TFBS than those of other hominids D4ex3 VNTRs studied (e.g. *P. troglodytes*, *Gorilla* sp or *P. pygmaeus*). Furthermore, the TFBSs in the *H. sapiens* D4ex3 VNTRs were similar to the hypothetical hominid ancestor VNTR. This suggest that, regardless of being located in a exonic region, the turnover of the TFBS of the D4ex3 VNTR sequences was not similar to that of non-coding ECR sequences evolving by genetic drift (e.g. Figure 3.13b) and this turnover rate was similar to that exhibit by fast evolving regulatory domains (e.g. the STin2 VNTR, Figure 3.10a).

In conclusion, I have shown that during evolution, the VNTRs in the DRD4 and SLC6A4 genes of *H. sapiens* and closest living relatives the great apes have diverged (Figure 3.22). As displayed in this diagram, the TFBSs of these VNTRs vary greatly amongst hominids, but importantly, these were also always markedly different

between *H. sapiens* and *P. troglodytes*. This is of special interest since, one of the aims of this thesis were to identify differences in the *cis* regulators of the expression of genes associated with modern human cognition that may have appeared as a response mechanism to the adaptation of the brain of *H. sapiens* during evolution.

Comparison of the *cis* regulatory elements of *H. sapiens* and the great apes allowed the identification of such potential changes. Therefore, the functional differences that may arise from variation in their VNTR sequences could correlate with differential *cis* regulation of the expression of the DRD4 and SLC6A4 genes *in vitro* and *in vivo*. To investigate this potential functional differentiation I conducted functional assays of the VNTRs transcriptional activities *in vitro* in different cell culture models, which are described in chapter 4.

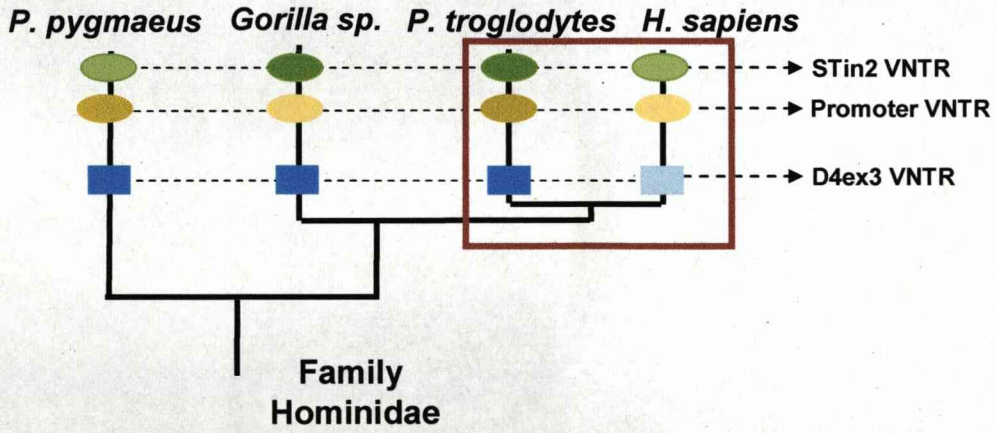


Figure 3.22 Diversification of the TFBS in the STin2, promoter VNTRs (SLC6A4 gene) and in the D4ex3 VNTRs (DRD4 gene) of hominids. The differences in TFBS found in the D4ex3, STin2 and promoter VNTRs of homindis are represented by geometric forms of varying shades (based on Figures 3.9, 3.12 and 3.20). For example, the types of TFBS the STin2 VNTRs found in hominids are similar in *P. troglodytes* and *Gorilla* sp, but markedly different from the TFBS found in *P. pygmaeus* and *H. sapiens*. Thus, the ovals representing the STin2 VNTRs for the first two species are dark green ovals and the VNTRs of the latter two are represented by light green ovals. Notice that in the tree, the shades of the shapes representing the promoter, D4ex3 and STin2 VNTRs of *H. sapiens* and *P. troglodytes* are different (marked by red rectangle).

Chapter 4. VNTRs in the DRD4 and SLC6A4 genes can support intra and inter-specific levels of reporter gene expression

4.1 Introduction

The copy number and sequence variability of the VNTRs in the SLC6A4 (STin2 and 5'promoter) and DRD4 (D4ex3) genes of *H. sapiens* have been correlated with differential transcriptional activity *in vitro* (Heils et al., 1996; Schoots and Van Tol, 2003; Lovejoy et al., 2003; Michellaugh et al., 2001; Klenova et al., 2005; Roberts et al., 2007) and *in vivo* (Mackenzie and Quinn, 1999). In chapter 3, the analysis of the sequences of these VNTRs in the SLC6A4 and DRD4 genes of primates demonstrated that in these species, the VNTRs exhibit great intra and inter-specific variability which affect the presence of putative TFBS (e.g. Table 3.1 and 3.2). The sequence analysis further revealed the existence of variation between the STin2 and 5' promoter VNTRs (SLC6A4 gene) and the D4ex3 VNTRs (DRD4 gene) of modern human and great apes. These results suggested that the VNTRs in the DRD4 and SLC6A4 genes of non-human primates might also act as tissue specific transcriptional domains and contribute to the intra-specific diversification of serotonin and dopamine neurotransmission in the brain as proposed for the *H. sapiens* VNTRs . Furthermore, the sequence variation found amongst hominid VNTRs suggests that their VNTR transcriptional activities may also be different. Therefore, in this chapter I analysed the potential *cis* regulatory properties of the VNTRs in different cell types, and their potential role in the intra and inter-specific diversity of gene expression.

4.2 Aims

The first aim of this chapter was to demonstrate that both the STin2 and the promoter VNTRs of the SLC6A4 and the D4ex3 VNTR of the DRD4 genes of non-human primates exhibit *cis* regulatory properties *in vitro*. For this study, I cloned

diverse primate VNTRs into pGL3p, a reporter gene expression vector (appendix 7) and tested their ability to support reporter gene expression in neuronal (dissociated cultures of neonate rat cortex and SN4741 clonal cells) and non-neuronal (JAr cells) cell culture models. The dissociated cultures of neonate rat frontal cortex were considered appropriate models for the study of the activity of these VNTRs in neurons because DRD4 and SLC6A4 genes were found endogenously expressed in this region (appendix 2 and McQueen et al., 1999; Araki et al., 2007). The human placental cell line JAr and the murine substantia nigra SN4741 cell line were considered useful models for the analysis of the VNTRs because they have been previously used in the analysis and characterization of activity of *cis* regulatory elements in the serotonergic and dopaminergic systems (Heils et al., 1995; Michelhaugh et al., 2001; Roberts et al., 2007; Sacchetti et al., 2001).

The second aim of this chapter was to analyse the possible correlation between intra-specific variation of the humans and non-human primates VNTR sequences and diversification of gene expression. For this study, I compared the transcriptional activities of copy variants of the VNTRs found in the SLC6A4 genes of *H. sapiens* and *P. troglodytes* when transfected into JAr and dissociated cultures of neonate rat cortex.

The third aim was to investigate the possible correlation between the inter-specific differences observed between the SLC6A4 and DRD4 VNTRs sequences of primates and their capacity to support differential levels of transcriptional activity *in vitro*. Therefore, I compared the transcriptional activities of constructs bearing the VNTR of different primate species (of the SLC6A4 and DRD4 genes) *in vitro* in dissociated cultures of neonate rat cortex and in SN4741 cells (D4ex3 VNTR only).

		SLC6A4 gene			
		STin2 VNTR		promoter VNTR	
Family	Species	No of repeat units	Name	No of repeat units	Name
Hominidae	*Homo sapiens*	9	STin2.9	14	Homo 14
		10	STin2.10	16	Homo 16
		12	STin2.12	–	–
	Pan troglodytes	18	STin2.18	17.5	Pan17.5
		19	STin2.19	–	–
		23	STin2.23	–	–
	Gorilla sp	40	STin2.40	18	Gor 18
	Pongo pygmaeus	6	STin2.6	20	Pon20
Cercopithecidae	*Cercopithecus aethiops*	5	STin2.5ca	–	–
	Mandrillus sphinx	5	STin2.5ms	–	–

Table 4.1 STin2 and promoter VNTR constructs used in this chapter. VNTRs from primates included in the families Hominidae and Cercopithecidae were cloned into luciferase expression vectors (pGL3p) and transfected into different cell models. In the graph, spacer lines introduced when no VNTRs constructs from one species were used.

4.3 Results

4.3.1 Correlation of the VNTR sequence variability to diversification of *cis* regulation of the SLC6A4 gene in primates

4.3.1.1 STin2 VNTR alleles of the SCL6A4 gene of *H. sapiens* support differential reporter gene expression in neuronal cultures and JAr cells

Variability of the VNTR in the second intron of the SLC6A4 gene has been implicated in the regulation of SLC6A4 expression in *H. sapiens*. The *cis* regulatory properties of the 3 most common alleles of the *H. sapiens* VNTR (STin2.9, STin2.10 and STin2.12) have been demonstrated *in vitro* in several clonal cell lines (Lovejoy et al., 2003, Klenova, et al, 2004). The *in vivo* transcriptional activities of constructs bearing the VNTRs with 10 and 12 repeats have been demonstrated *in vivo* in a transgenic mouse embryo model (Mackenzie and Quinn, 1999); however, the Stin2.9 construct was not tested. In spite of the potential role of the STin2 VNTR in mediating differential SLC6A4 expression in the brain of *H. sapiens*, the putative differential transcriptional activities of these 3 copy number variants have not been compared in neuronal cultures. Therefore, I analysed the transcriptional activity of the three STin2 VNTR variants of *H. sapiens* SLC6A4 gene and in dissociated cultures of neonate rat frontal cortex and for comparative reasons they were also transfected into JAr cells (described in Table 4.1). Sections of frontal cortex were obtained from 2-7 day old male wistar rats and dissociated as described in section 2.2.15.1. Transfections were conducted following the protocol described in section 2.2.16. The basal transcriptional activities were calculated as detailed in section 2.2.18.1. The transfection efficiency of the constructs was standardised following protocol in section 2.2.19.1. Results are presented in Figure 4.1 and Table 4.2.

Transfection of the VNTR constructs into JAr cells (Figure 4.1) confirmed that the 3 STin2 VNTRs of *H. sapiens* are capable of supporting reporter gene expression different from pGL3p (Table 4.2). In this study, the STin2.9 supported strongest expression followed by STin2.12 and STin2.10 (Table 4.2); however, these differences were not statistically significant.

In the primary cultures of neonate rat cortex, the STin2.9 construct supported higher levels of reporter gene expression than STin2.12 and STin2.10 (Figure 4.1). Compared to their transcriptional activities in JAr cells, the levels of activity supported by the STin2.9 and STin2.12 showed a significant increase in cortical cultures (1.6 and 1.4 fold increase respectively, *T*-test p=0.05, Figure 4.1); conversely, the activity of the STin2.10 construct did not change significantly. In both models the average reported gene expression supported by the three constructs was different, but only in the cortical cultures these differences were statistically significant, with the STin2.10 construct supporting lower levels of reporter gene expression than either STin2.12 or STin2.9 (Table 4.2).

The results suggest that the three most commonly found variants of the STin2 VNTR of *H. sapiens* exhibit differential *cis* regulatory properties in neuronal tissue, and this may corelate to the association proposed to exist between the STin2 genotype and serotonin related behaviour in the modern human population (Kremer et al., 2005; Mulder et al., 2005; Payton et al., 2005). The study would be consistent with the activities of these STin2 VNTRs being variable in different tissues, based on the differential activity of reporter gene supported by the 3 different constructs in the two *in vitro* models.

Figure 4.1

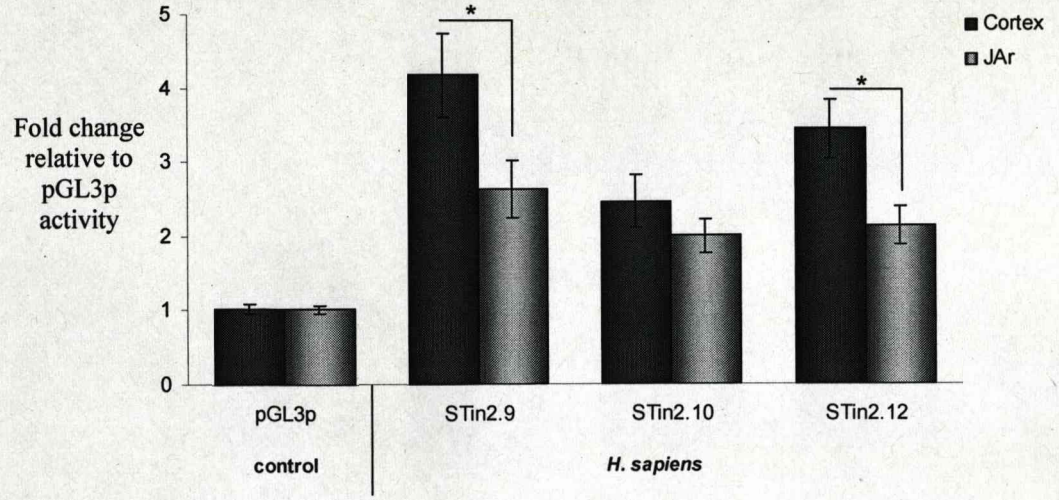


Table 4.2

	Construct name	pGL3p	STin2.9	STin2.10	STin2.12
Cortex	average fold increase over pGL3p	1.000	4.161	2.454** ¶	3.426
		±	±	±	±
	standard error	0.063	0.568	0.352	0.392
	Construct name	**pGL3p**	**STin2.9**	**STin2.10**	**STin2.12**
Jar	average fold increase over pGL3p	1.000	2.611	1.989	2.120
		±	±	±	±
	standard error	0.061	0.387	0.231	0.253

Figure 4.1. The STin2 VNTRs of the SLC6A4 gene of *H. sapiens* supported reporter gene expression in JAr cells and primary cultures of rat cortex. STin2 VNTRs and pGL3p (control) were transfected (1μg each) into JAr cells and into dissociated cultures of neonate rat frontal cortex using TRANSFAST and ExGen 500 respectively. Student's *T*-test, showed levels of reporter gene expression supported by STin2.9 and STin2.12 were significantly higher in cortical cultures than in JAr cells (indicated by *, Student's *T*-test $p \leq 0.05$). The activities supported by the STin2 VNTR constructs in cortical cultures and JAr cells are represented by dark grey and light grey bars respectively. Error bars are based on a minimum of 3 independent experiments (in triplicate wells, $n=9$).

Table 4.2 The three VNTR constructs supported differential reporter gene expression only when transfected into cortical cultures. The average activities of the 3 STin2 VNTRs increase when transfected into cortical cultures. In particular, the activities of STin2.9 and STin2.12 were significantly higher than that of STin2.10 (where ** and ¶ indicated significant differences between the activties of STin2.10 and STin2.9 and STin2.12 respectively).

4.3.1.2 STin2 VNTR of the SLC6A4 gene of *P. troglodytes* exhibit differential transcriptional activities.

The STin2 VNTR SLC6A4 gene of *P. troglodytes* presents abundant copy number and sequence variation (Figure 3.7). Analysis of the sequence of the STin2 VNTRs of *P. troglodytes* shows that such variability produces diversification in the number and type of putative TFBS present in the different variants (appendix 4.1). This variability could contribute to the diversification of SLC6A4 expression in the *P. troglodytes* population, as proposed for the *H. sapiens* STin2 VNTRs. To asses if the STin2 VNTR locus in *P. troglodytes* could be associated with differential regulation of the SLC6A4 gene, I cloned the three common copy number variants (STin2.18, STin2.19 and STin2.23) found in (Figure 3.7) into pGL3p to analyse their *in vitro* transcriptional activities. The 3 constructs were transfected in JAr and dissociated cultures of neonate rat frontal cortex and transfected following protocols described in section 2.2.16.1 and 2.2.16.3. Frontal cortex sections were obtained from neonate wistar rats aged 2-7 days. Analysis of the reporter gene expression was assessed as described in section 2.2.18.1.

In JAr cells (Figure 4.2), all three STin2 VNTRs of *P. troglodytes* supported levels of reporter gene expression that were only marginally different from pGL3p (Table 4.3). The STin2.18 VNTR construct supported higher levels of reporter gene expression than STin2.23, which was as active as pGL3p alone. When the three VNTRs constructs were transfected into the cortical cultures, the three STin2 VNTRs were now capable of driving reporter gene expression (Figure 4.2). The average transcriptional activity of all 3 constructs increased, as observed for the STin2 VNTR constructs of *H. sapiens,* being significantly higher for the STin2.19 and STin2.23

constructs (Table 4.3). The results presented here constitute the first report of *cis* regulatory activities of the STin2 VNTR of the SLC6A4 gene of *P. troglodytes*.

These findings suggest that the STin2 VNTRs of the SLC6A4 gene of *P. troglodytes* may contribute to the regulation of the SLC6A4 expression in the CNS, but their transcriptional activities are distinct from those exibit by the STin2 VNTRs of *H. sapiens*. However, their capacities to generate differential regulation of gene expression remain to be demonstrated, as the average transcriptional activities of the constructs in cortex cell cultures were very similar. Nevertheless, the potential contribution of these STin2 VNTRs to diversification of serotonin related behaviour in the population of *P. troglodytes* might only become apparent when an appropriate stress or challenge is applied. This is corroborated by the significant activation of the constructs only when transfected into cortical cultures. It is possible that the differences in the transcriptional activities of the constructs be more readily identified using a different experimental set up, such as transgenic models, previously used for analysing the effect of sequence variation in the STin2 VNTRs *cis* regulatory properties (Mackenzie and Quinn, 1999).

Figure 4.2

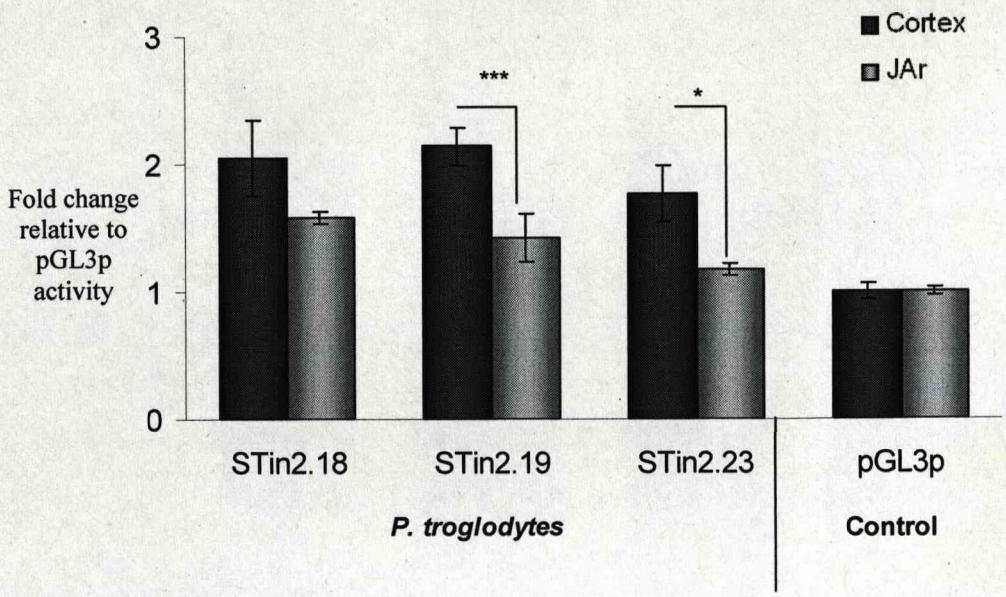


Table 4.3

		Construct name			
		pGL3p	**STin2.18**	**STin2.19**	**STin2.23**
JAr	average fold increase over pGL3p	1.000	1.578	1.423	1.167
		±	±	±	±
	standard error	0.033	0.049	0.194	0.047
		pGL3p	**STin2.18**	**STin2.19**	**STin2.23**
Cortex	average fold increase over pGL3p	1.000	2.045	2.140	1.765
		±	±	±	±
	standard error	0.061	0.293	0.148	0.219

Figure 4.2. The STin2 VNTRs of the SLC6A4 gene of *P. troglodytes* supported reporter gene expression in JAr cells and primary cultures of rat cortex. STin2 VNTRs and pGL3p (control) were transfected (1 μg each) into JAr cells and into dissociated cultures of neonate rat frontal cortex using TRANSFAST and ExGen 500 respectively. The STin2 VNTRs activities in cortical cultures are represented by dark grey bars and their activities in JAr cells are represented by a light grey bars. The activities of the VNTR constructs was marginal in JAr cells whereas they significantly increase in cortical cultures (Student's *T*-test * = $p \leq 0.05$, and *** = $p < 0.001$).

Table 4.3. Differential transcriptional activities of STin2.19 and STin2.23 VNTR constructs in the two *in vitro* models. The *P. troglodytes* STin2 VNTR constructs supported marginal reporter gene expression when transfected into JAr cultures but the activities of all 3 constructs increase when transfected into cortical cultures.

4.3.1.3 The promoter VNTRs of the SLC6A4 gene of *H. sapiens* support differential levels of reporter gene expression in dissociated cultures of rat neonate frontal cortex

In modern humans, the promoter VNTR of the SLC6A4 gene presents copy number variants with 14 to 20 repeats units and several SNPs. The variants with 14 and 16 repeat units (a.k.a. 5HTTLPR-short and 5HTTLPR-long) are the most prevalent in all modern human populations studied to date (Gerlenter et al., 1999). These two variants differ in their copy number, and this difference has been linked to the different *in vitro* transcriptional profiles (Heils et al., 1996) and to differential expression of the SLC6A4 in the human brain *in vivo* (Hariri et al., 2002b; Hranilovic et al., 2004). However, their ability to support reporter gene expression in neurons or tissue derived from CNS has not been demonstrated to date. To investigate this, I transfected *H. sapiens* constructs bearing the VNTRs with 14 and 16 repeats (generated by F. Ali and A. Sharda) cloned into pGL3p and delivered them into dissociated cultures of neonate rat frontal cortex (Figure 4.3).

The present study shows that the Homo 16 and Homo 14 promoter VNTRs of the SLC6A4 gene of *H. sapiens* are capable of supporting differential reporter gene expression in neuronal cultures (≈3.5 fold and ≈7.6 fold increase of the Homo 16 and Homo 14 respectively, Figure 4.3). Although absolute levels of reporter gene expression varied possibly caused by heterogeneity in primary cultures Homo 14 VNTR consistently supported 2.2 more luciferase reporter gene expression than Homo 16 VNTR, and this difference was significant (Student's *T*-test p=0.048) in all experiments. The present study suggests that the ratio of the activities of these two VNTR variants described in cell lines remains constant in a neuronal environment. Furthermore, the difference in the *cis* regulatory activities of the Homo 14 and Homo

16 VNTRs seen in the cortical cultures correlates with the results of association studies, which have linked to the promoter VNTRs genotype to intra-specific diversification of SLC6A4 gene expression modern humans.

Interestingly, in this study the two VNTR constructs acted as enhancers, and not as silencers of reporter gene expression as reported previously (Lesch et al., 1997; Sakai et al., 2002). The discrepancies between the activity of promoter VNTRs of the SLC6A4 gene of *H. sapiens* found in the present study and studies conducted by others in different *in vitro* models are likely to reflect the tissue specificity of the these promoter VNTRs. It is worth noting that in such previous studies the two VNTR constructs supported transcriptional activities which were different in 2.5- 2.7 fold, similar as found in the present report. However, in such studies the construct bearing the long and not the short allele of the promoter VNTR supported the highest levels of reporter gene expression (Heils et al., 1997). I propose that the studies in the primary cultures derived from the CNS are more relevant for assessing real functional differences of this VNTR than those conducted in cell lines such as JAr limphoblastma cells lines.

Figure 4.3

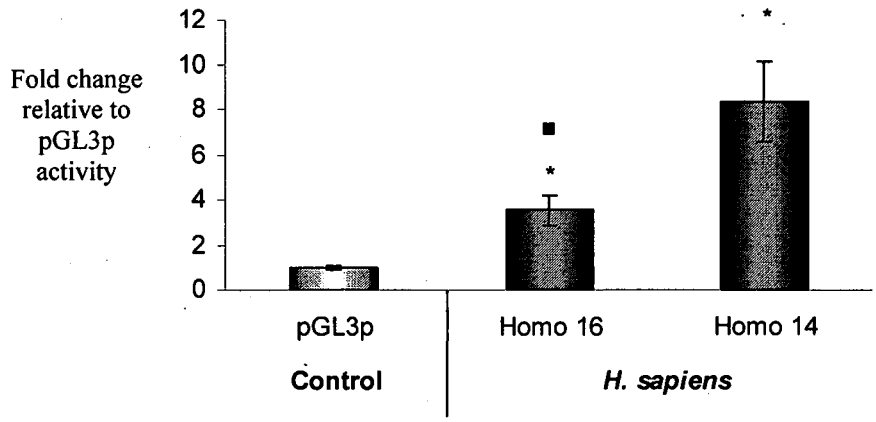


Figure 4.3 The promoter VNTRs of the SLC6A4 gene of *H. sapiens* supported differential levels of reporter gene expression in neuronal cultures. Promoter VNTRs and pGL3p (control) were transfected (1μg each) into dissociated cultures of neonate rat frontal cortex using ExGen 500 (Activities supported by the promoter VNTR constructs are represented by dark grey and that of pGL3p is represented by a light grey bar). Student's *T*-test, showed that both VNTR constructs supported reporter gene expression different from pGL3p (where *= $p \leq 0.05$). The transcriptional activities of the two VNTR constructs were signifcantly different (where ■= $p \leq 0.05$). Error bars are based on values from a minimum of 3 independent experiments (in triplicate wells; n=9).

4.3.2 Correlation between DRD4 and SLC6A4 VNTRs sequence evolution and their transcriptional activities *in vitro*

4.3.2.1 Divergence between the sequences of *H. sapiens*, *P. troglodytes* and *Gorilla* sp. STin2 VNTRs parallel the VNTR transcriptional activities in cortical cultures.

In chapter 3, I compared the putative TFBS found in the STin2 VNTRs of hominids and their evolution. This analysis suggested that during evolution, the sequences of African apes *P. troglodytes* and *Gorilla* sp STin2 VNTR had undergone great differentiation from the sequence of the hypothetical ancestral hominid, and from the STin2 VNTR sequences of *H. sapiens* and *P. pygmaeus*.(Figure 3.7 and 3.10a). In the previous section I demonstrated that sequence variation of VNTRs in the *H. sapiens* SLC6A4 gene correlates with differential transcriptional activities *in vitro*. Similarly, the difference amongst hominids STin2 VNTRs suggests potential inter-specific differences in *cis* regulation of the SLC6A4 gene. To test this hypothesis, I compared the ability of the STin2 VNTRs of the SLC6A4 gene of *H. sapiens* and African great apes (*P. troglodytes*, *Gorilla* sp.) to support reporter gene expression in neuronal cultures. The transfection of constructs, preparation of cell cultures and measurement of transcriptional activities of the VNTR constructs have been described before in section 4.3.1.1.

In primary cultures of neonate rat cortex, the levels of luciferase gene expression supported by the STin2 VNTRs constructs of African hominids (Table 4.1) showed significant differences from the transcriptional activities supported by the *H. sapiens* VNTR constructs (Figure 4.4). The STin2 VNTR construct of *Gorilla* sp. did not produced levels of reporter gene expression (Table 4.4). The three STin2 VNTRs of *P. troglodytes* (STin2.18, STin2.19 and STin2.23) supported levels of reporter gene

expression lower than *H. sapiens* STin2.9 and STin2.12 but not from STin2.10 (Table 4.4).

The results of these functional assays suggests that although there is an overlap between the transcriptional activities of STin2.10 of *H. sapiens* and the VNTR constructs of *P. troglodytes*, there is a trend where the regulatory activities of the *H. sapiens* STin2 VNTRs to be higher than those of African apes VNTRs in this model. In consequence, this distinction could contribute to the differentiation of the *cis* regulation of the SLC6A4 gene amongst these species in the CNS.

Figure 4.4

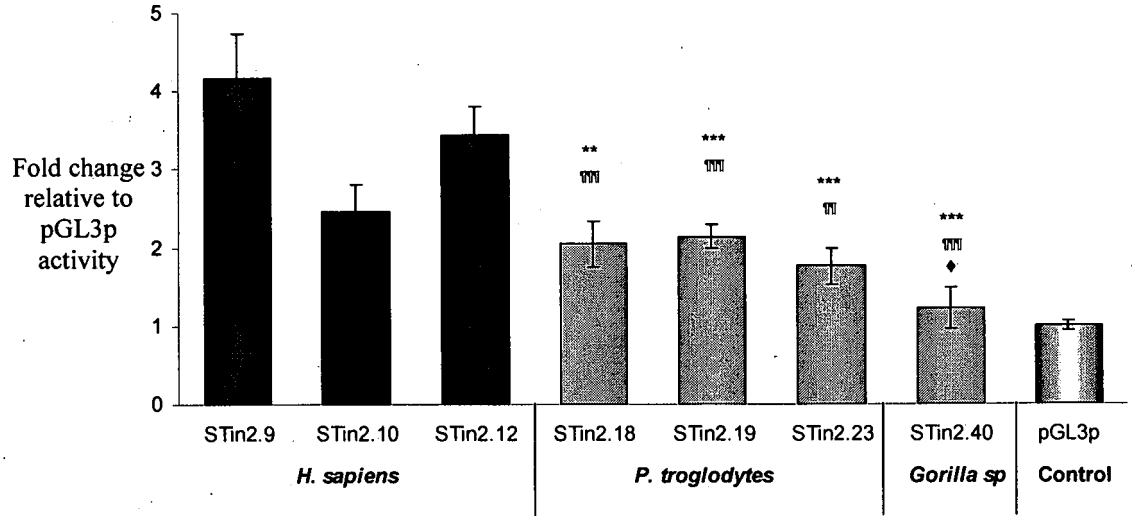


Table 4.4

	H. sapiens			P. troglodytes			Gorilla sp	control
Construct name	**STin2.9**	**STin2.10**	**STin2.12**	**STin2.18**	**STin2.19**	**STin2.23**	**STin2.40**	**pGL3p**
Average fold increase over pGL3p standard error	4.161 ± 0.568	2.452 ± 0.352	3.426 ± 0.367	2.045 ± 0.293	2.140 ± 0.148	1.765 ± 0.232	1.219 ± 0.264	1.000 ± 0.061

Figure 4.4. Comparison of the levels of reporter gene expression supported by the STin2 VNTRs of the SLC6A4 gene of hominids in neuronal cultures. STin2 VNTRs constructs were transfected into dissociated cultures of neonate rat frontal cortex using ExGen 500. Dark grey bars represent activities of *H. sapiens* VNTR constructs whereas the light grey bars represent those of the *P. troglodytes* and *Gorilla* sp. STin2 VNTR constructs. ** and *** indicate very significant and very highly significant differences ($p<0.01$ and $p<0.005$) between the levels of reporter gene expression of the great ape STin2 VNTRs vs. *H. sapiens* STin2.9; ¶¶ and ¶¶¶ indicate significant differences of activities of great ape STin2 VNTRs and that of STin2.12.

Table 4.4. The *H. sapiens*, *P. troglodytes* and *Gorilla* sp VNTR constructs support differential levels of reporter gene expression *in vitro*. Student's *T*-test showed significant differences between the transcriptional activities supported by the pGL3p and the STin2 VNTR constructs. Amongst all constructs tested the STin2.40 (*Gorilla* sp.) supported the lowest levels of reporter gene expression, which were not distinct from those of pGL3p.

4.3.2.2 Similarity in the *H. sapiens* and *P. pygmaeus* STin2 VNTR sequences correlates with *in vitro* ability to support reporter gene expression

The analysis of the STin2 VNTR sequence diversity and evolution suggested that the STin2 VNTRs of *H. sapiens* and *P. pygmaeus* did not differ greatly from the VNTR of the ancestral hominid (Figure 3.10a), and consequently their sequences shared some putative TFBSs. To investigate if the commonalities found in the sequences of *H. sapiens* and *P. pygmaeus* STin2 VNTRs could correlate with similarity in their VNTR transcriptional activities, I compared their supported transcriptional activities when transfected into dissociated cultures of neonate rat frontal cortex.

The results are shown in Figure 4.5. In brief, this study shows that the STin2 VNTR in *P. pygmaeus* posses transcriptional activity in cortical cultures which are similar to that supported by *H. sapiens* VNTRs. Thus, this VNTR may act as a *cis* regulator of the SLC6A4 gene expression in this species. Statistical analysis showed that the activities of the STin2 VNTRs of *H. sapiens* and *P. pygmaeus* were not significantly different (Figure 4.5, Student's *T*-test, p=0.98, 0.1 and 0.46 respectively) in this cell culture model.

The results of this preliminary study suggest that the few changes accumulated by the STin2 VNTRs of *H. sapiens* and *P. pygmaeus* since they last shared a common ancestor may correlate with similar transcriptional properties. However, to corroborate this hypothesis, a bigger sample size of *P. pygmaeus* and *H. sapiens* than the used for this study must be included to ensure that allelic diversity known to affect the transcriptional activities of these VNTRs is considered.

Figure 4.5

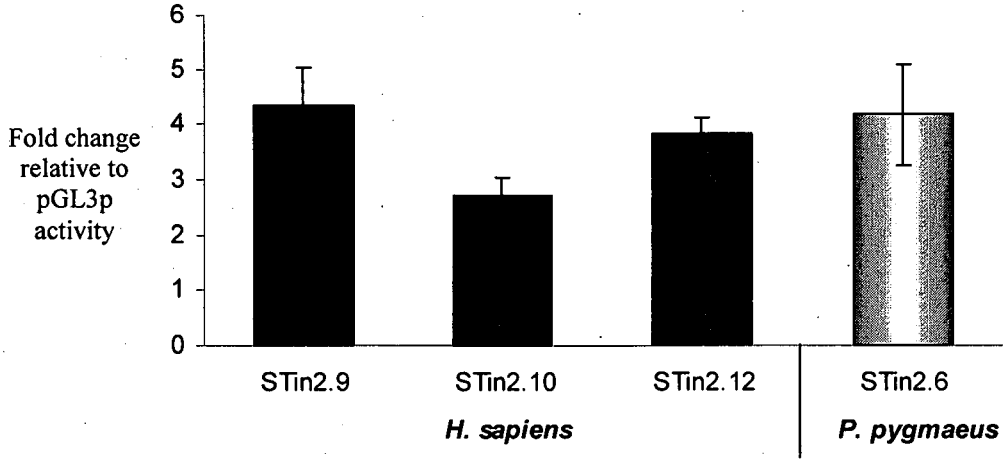


Figure 4.5 Comparison of the reporter gene expression levels supported by the STin2 VNTRs of the SLC6A4 gene of *H. sapiens* and *P. pygmaeus* in neuronal cultures. The STin2 VNTRs constructs were transfected into dissociated cultures of neonate rat frontal cortex using ExGen 500. In this cell model the activities of the *P. pygmaeus* and *H. sapiens* STin2 VNTR constructs supported similar levels of reporter gene expression. Transcriptional activities of the constructs were calculated as the fold change of the activity supported by the VNTR constructs relative to pGL3p activity. Activities of *H. sapiens* STin2 VNTR constructs are represented by black bars and that of *P. pygmaeus* STin2 VNTR construct is represented by a light grey bar. Error bars are based on values from a minimum of 3 independent experiments (in triplicate wells; n=9).

4.3.2.3 The STin2 VNTRs of old world monkeys support higher levels of reporter gene expression in dissociated cultures of neonate rat frontal cortex than in JAr cells

The ancestral taxa that gave rise to modern apes (family Hominidae) and old world monkeys (family Cercopithecidae) last shared a common ancestor 25-35 mya (Page and Goodman 2001). In spite of the time passed, the SLC6A4 gene old world monkeys present a STin2 VNTR, which is orthologous to the VNTR present in the SLC6A4 gene of hominids (Figure 3.8). Analysis of the sequences of the cercopithecids STin2 VNTR showed that these VNTRs present abundant TFBS; suggesting that the SLC6A4 gene expression of cercopithecids could also be modulated via the STin2 VNTR. To address whether the STin2 VNTRs of cercopithecids exhibit *cis* regulatory properties as exhibited by those of hominids, I cloned *Cercopithecus aethiops* (green vervet) and *Mandrillus sphinx* (mandrill) SLC6A4 STin2 VNTRs into pGL3p and transfected them into JAr cells and into dissociated cultures of rat neonate cortex.

Transfection of the two constructs in JAr cells (Figure 4.6) demonstrates that both STin2 VNTRs were able to support low reporter gene expression, but significantly different from pGL3 alone (Figure 4.6, Table 4.5). In this *in vitro* model, the levels of transcriptional activities supported by both constructs were not significantly different (Student's *T*-test p=0.108). In dissociated cultures of neonate rat cortex (Figure 4.6), both STin2 VNTR constructs were capable of supporting reporter gene expression (6.01 and 11.743 average fold of STin2.5ca and STin2.5ms respectively) and their activities were significantly higher than the observed in JAr cells (Table 4.5). In this model, the average transcriptional activities of the two STin2

VNTR constructs were significantly different (*M. sphinx* VNTR construct was 2 fold more active than *C. aethiops* VNTR; Table 4.5; Student's *T*-test, p=0.02)

This study suggests that the *cis* regulatory capacity of the STin2 VNTRs of the SLC6A4 of primates arise at least 25-35 mya, before Cercopithecidae and Hominidae families separated. Moreover, the *in vitro* assays presented here demonstrate that the activities of the STin2 VNTRs are tissue specific and perhaps contribute to the differentiation of SLC6A4 gene expression amongst cercopithecids.

Figure 4.6

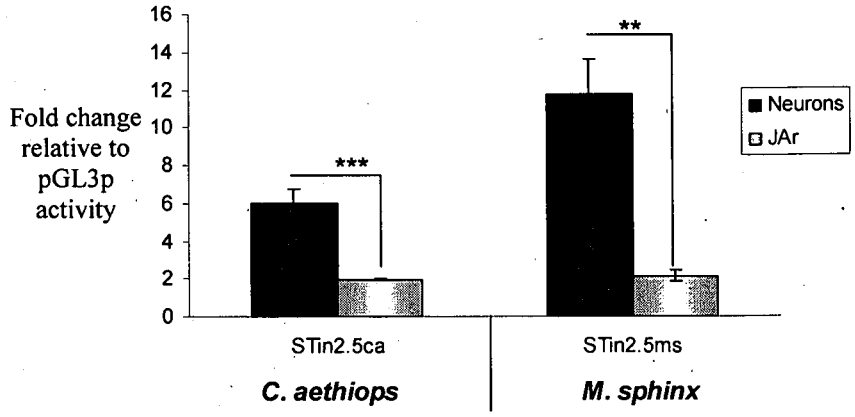


Table 4.5

	Cortical cultures		JAr cells		Cortex/JAr	
Fold increase over pGL3p	STin2.5ca	STin2.5ms	STin2.5ca	STin2.5ms	STin2.5ca	STin2.5ms
	6.012 ■ ***	11.743 ***	1.946 ***	2.133 ***	3.089 ***	5.505 **
	±	±	±	±	±	±
SE	0.393	2.058	0.035	0.300	0.26	0.970

Figure 4.6 The transcriptional activities of STin2 VNTRs of the SLC6A4 gene of *C. aethiops* (STin2.5ca) and *M. sphinx* (STin2.5ms) in neuronal cultures and JAr cells. VNTRs constructs (1 μg each) were transfected into JAr cells and cortical cultures using TRANSFAST and ExGen 500 respectively. The activities of the constructs significantly increase when transfected into cortical cultures (Student's *T*-test, showed significant differences in transcriptional activities; ** =p≤0.01 and *** =p≤0.001). Activities of constructs in neuronal and JAr cultures were represented by black bars and light grey bars respectively.

Table 4.5 STin2.5ms and STin2.5ca supported differential reporter gene expression in cortical cultures. In both models the VNTR constructs supported reporter gene expression different from pGL3p (as indicated by ***, Student's *T*-test p≤0.005). Student's *T*-test demonstrated significant differences between the levels of transcriptional activities supported by the two constructs in cortical cultures (where ■ indicates p≤0.05). The levels of activity of the constructs significantly increase when transfected into cortical cultures (indicated by ** =p≤0.01 and *** =p≤0.001).

4.3.2.4 Cercopithecids and *P. pygmaeus* STin2 VNTR sequence similarity correlates with their transcriptional activities dissociated cultures of neonate rat cortex

The lineage from which *P. pygmaeus* originated was the first to separate from the ancestral ape stock. As such, it is likely that this species resemble more genetically and phenotypically the ancestral apes than any of the other extant hominids (genera *Homo*, *Pan* and *Gorilla*). This hypothesis is corroborated by similarities shared between the sequences of this cercopithecids and *P. troglodytes* (as represented in Figure 3.10a), and in addition, cercopithecids are often used as reference for early hominids. To test whether the similarities between the sequences of cercopithecids and *P. troglodytes* STin2 VNTRs correlate with their transcriptional activities *in vitro*, I transfected the luciferase constructs bearing their VNTRs into dissociated cultures of neonate rat cortex (Figure 4.7).

The functional assay demonstrated that the STin2 VNTR of *P. pygmaeus* is capable of supporting reporter gene expression *in vitro* (4.18 fold increase, Figure 4.7). Furthermore, student's *T*-test demonstrated that the transcriptional activities of the STin2 VNTRs of *P. pygmaeus* and cercopithecids *M. sphinx* and *C. aethiops* were not significantly different in the cortical cultures (STin2.6 vs. STin2.5ca $p=0.83$ and STin2.6 vs. STin2.5ms $p=0.1$, Figure 4.7). This result shows that as predicted by the sequence similarity identified in chapter 3 (Figure 3.10a), the transcriptional activities of *P. pygmaeus* STin2 VNTR resemble those of the cercopithecids STin2 VNTRs. Moreover, suggest that the STin2 of the ancestral apes would have been similar to the transcriptional activities exhibited by these *P. pygmaeus* and cercopithecids STin2 VNTRs.

Figure 4.7

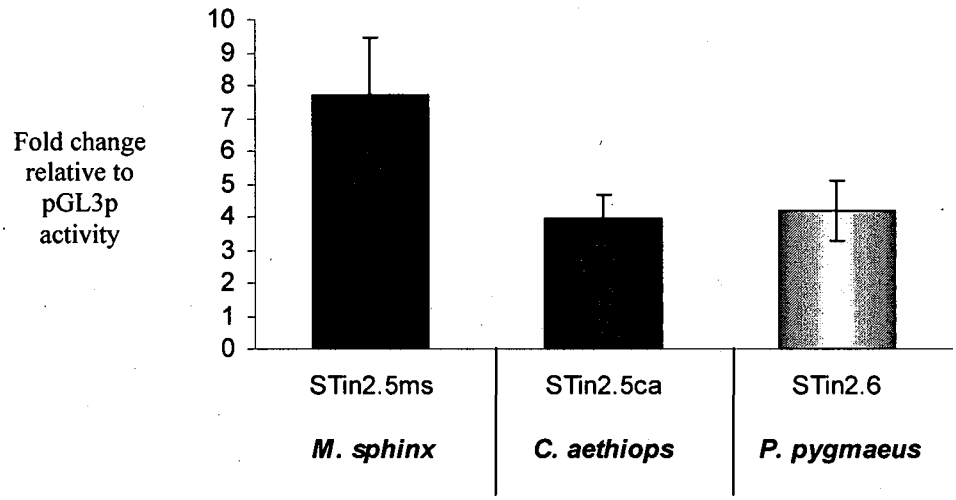


Figure 4.7 Comparison of the transcriptional activities of the STin2 VNTRs of cercopithecids and *P. pygmaeus*. The activities of the cercopithecids and *P. pygmaeus* STin2 VNTR constructs supported similar levels of reporter gene expression in cortical cultures. Transcriptional activities of the constructs were calculated as the fold change of the activity supported by the VNTR constructs relative to pGL3p activity. Activities of cercopithecids STin2 VNTRs are represented by black bars and that of *P. pygmaeus* STin2 VNTR construct is represented by a light grey bar. Error bars are based on values from a minimum of 3 independent experiments (in triplicate wells; n=9).

4.3.2.5 Comparison of the *cis* regulatory abilities of the promoter VNTR of the SLC6A4 genes of hominids

In section 4.3.1.3, it was demonstrated that the promoter VNTRs of the *H. sapiens* are capable to support differential reporter gene expression in cortical cultures. Analysis of the evolution of the promoter VNTR sequences of SLC6A4 gene of hominids (Figure 3.13a) showed that these have not diverged as much as other VNTRs studied (STin2 and DRD4) have done. Nevertheless, the 5' promoter VNTR sequences present traits (TFBS and repeat units) which distinguish each species variants (Figure 3.11). To analyse the effects of the sequence similarity in the transcriptional properties of the promoter VNTRs of hominids, I transfected the promoter VNTR constructs (described in Table 4.1) into dissociated cultures of neonate rat cortex and compared the supported levels of luciferase expression.

4.3.2.5.1 The transcriptional activities of the promoter VNTRs of *H. sapiens* is distinct from the *Gorilla* sp. promoter VNTR.

The analysis showed that *P. pygmaeus* and *Gorilla* sp. variants were capable of supporting reporter gene expression in neuronal environment (2.06 and 5.436 fold increase respectively over pGL3p, Figure 4.8). Given the sequence variation identified in both species (as shown in Figure 3.11) it is possible that the 5' promoter VNTR could act as a transcriptional regulatory domain as proposed for the *H. sapiens* and contribute to the intra-specific diversification of SLC6A4 expression in these species. Furthermore, the results of the comparison of *in vitro* reporter gene expression supported by the VNTRs of these species and by the variants of *H. sapiens* (Table 4.6), suggest a potential overlap of their transcriptional properties in a neuronal environment. Indeed, the levels of reporter gene expression supported by the *H.*

sapiens VNTRs (Homo 14 and Homo 16) were indistinguishable from the supported by the *Gorilla* sp. VNTR construct (Gor 18) (Figure 4.8). Although the average level of reporter gene expression supported by the promoter VNTR construct of *P. pygmaeus* was lower than the levels supported by both *H. sapiens* constructs (Figure 4.8); this difference was only significant between the Pon20 and Homo14 constructs (Table 4.6). Furthermore, the overlap in the activities of the promoter VNTRs of *H. sapiens*, *P. pygmaeus* and *Gorilla* sp. (Table 4.6) corroborates the possible correlation between the similarity in TFBS of the promoter VNTR (as represented in Figure 3.11) and their *cis* regulatory activity in neurons *in vitro*.

This analysis shows that small changes in the sequence of regulatory domains such as exhibited by the *H. sapiens* and *P. troglodytes* promoter VNTRs can reflect great differences in their transcriptional properties *in vitro*. The *P. troglodytes* promoter VNTR construct (Pt 17.5) did not support significant levels of reporter gene expression in the cortical cultures, and its activity was significantly different from the activities supported by both *H. sapiens* VNTR constructs (Table 4.6). It is possible that the Pt17.5 construct is transcriptionally active under cellular stress, the presence of certain stimuli or indeed in different cells. Nevertheless, this functional distinction suggests a functional divergence of the Pt17.5 construct from the activities of the promoter VNTRs in other hominids and importantly, from the activities of the *H. sapiens* VNTR constructs. These results are similar to the functional assay results of the hominids STin2 VNTRs (Figure 4.4), where the *H. sapiens* VNTRs were seen to support higher levels than the *P. troglodytes* constructs. In conclusion, these preliminary studies suggest potential differential *cis* regulation of the SLC6A4 gene of *H. sapiens* and *P. troglodytes* via the promoter and STin2 VNTRs. However, it is noteworthy that although the *P. troglodytes* variant used in this functional assay is

commonly found in this species, there are other variants reported (Lesch et al., 1997) which may exhibit high levels of transcriptional activity, and need to be addressed.

Figure 4.8

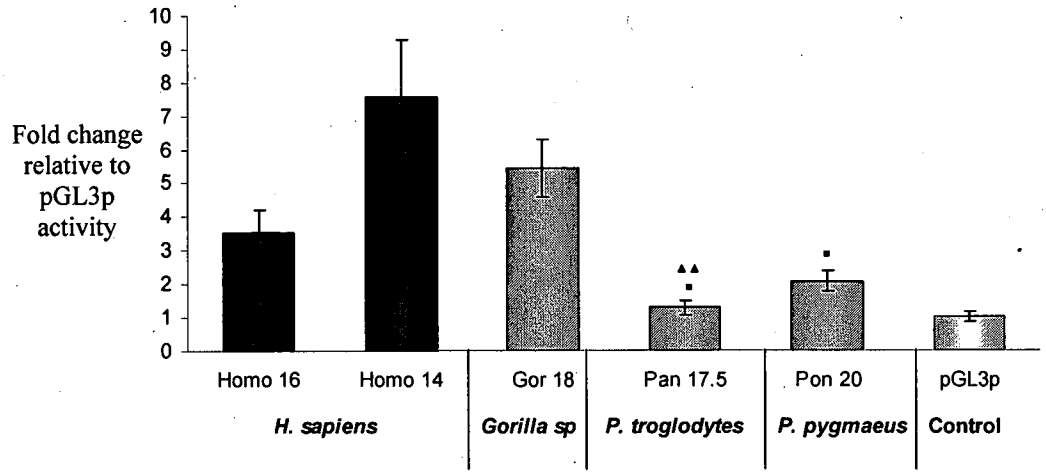


Table 4.6

	Homo 16	Homo 14	Pan 17.5	Gor 18	Pon 20	pGL3p
average fold increase	3.538	7.571	1.274	5.436**	2.060**	1.000
	±	±	±	±	±	±
standard error	0.678	1.704	0.233	0.862	0.312	0.125

Figure 4.8 Comparison of the reporter gene expression levels supported by the promoter VNTRs of hominids in cortical cultures. The transcriptional activities of the promoter VNTR constructs of *H. sapiens* were different to *P. troglodytes*, but overlapped with those of the *P. pygmaeus* and *Gorilla* promoter VNTRs constructs (Stundent's *T*-test, where ▲▲ and ■ indicate significat differences between the activity of a VNTR construct and the activities of Homo 14 and Homo 16 respectively). The constructs were transfected into dissociated cultures of neonate frontal cortex cultures, using ExGen 500. Transcriptional activities were calculated as the activities of the VNTR constructs relative to the pGL3p activity.

Table 4.6. The hominid promoter VNTR constructs support reporter gene expression *in vitro*. Student's *T*-test showed significant differences between the transcriptional activities supported by the great ape promoter constructs and pGL3p (**=$p \leq 0.01$; $n=8$).

4.3.2.6 The transcriptional activities of the D4ex3 VNTR of the DRD4 gene of *H. sapiens* and *P. troglodytes*

The VNTR in the third exon of the DRD4 gene (D4ex3 VNTR) presents great sequence variability in *H. sapiens* and other non-human primates (Table 3.2). *In vitro* assays have demonstrated that the three most commonly found variants of the D4ex3 VNTR of the modern human population are capable of supporting differential levels of reporter gene expression *in vitro* in the rat pituitary cell line GH4C1 (Schoots and Van Tol 2003). In spite of their potential role these elements play in the *cis* regulatory control of the DRD4 gene expression, their *cis* regulatory activities in neurons has not been demonstrated to date. Furthermore, the analysis of the primates D4ex3 VNTR sequences in chapter 3 showed that these VNTRs could potentially act as transcriptional domains in all primates. However, the differences between the TFBS in the *H. sapiens* and other hominids D4ex3 VNTR sequences (as represented in Figure 3.21) suggested that the *H. sapiens* VNTRs transcriptional activities might differ from that of the great apes VNTRs. To test these hypotheses, I conducted a preliminary study of the transcriptional properties of the D4ex3 VNTRs in *H. sapiens* and its closest living relative *P. troglodytes*.

For this study the most common VNTR variant found in *H. sapiens* (4 repeats) and a commonly found variant in *P. troglodytes* (5 repeats) were cloned into a luciferase reporter vector (pGL3p) and analysed in both SN4741 cells and dissociated cultures of rat frontal cortex. The results obtained are presented in Figure 4.9. In brief, only the *H. sapiens* construct was functional. In the SN4741 cells the *H. sapiens* construct (HD4ex3) supported significant levels of reporter gene expression (1.8 fold increase), whereas the VNTR construct of *P. troglodytes* (PtD4ex3) did not support

reporter gene expression (Table 4.7). The differences in the transcriptional activities supported by the two VNTR constructs were very highly significant (Table 4.7).

The activity of the *H. sapiens* VNTR construct reverses to a repressor in the dissociated cultures of rat neonate frontal cortex (Figure 4.9). In this model, the *H. sapiens* construct significantly repressed the activity of the pGL3p (in 55%; Student's *T*-test, $p<0.001$). The *P. troglodytes* VNTR construct did not exhibit transcriptional activity, as observed in SN4741 cells. The difference in the activity of the two constructs was very highly significant (Student's *T*-test $p< 0.001$, Table 4.7).

Table 4.7

		H. sapiens	*P. troglodytes*	control
		HD4ex3	**PtD4ex3**	**pGL3p**
Cortex	average fold increase over pGL3p	0.456 * ▪▪	0.793	1.001
		±	±	±
	standard error	0.034	0.049	0.040
		HD4ex3	**PtD4ex3**	**pGL3p**
SN4741	average fold increase over pGL3p	1.793 ** ▪▪▪	0.762	1.000
		±	±	±
	standard error	0.214	0.066	0.097

Table 4.7 The HD4ex3 supports significantly different reporter gene expression in cortical and in SN4741 cultures. Student's *T*-test shows that only HD4ex3 supported reporter gene expression when compared to the control (pGL3p) in both cell models (* and ** indicate $p\leq0.05$ and 0.01 respectively). The statistical test also showed significant differences between the levels of reporter gene expression supported by the two VNTR constructs (▪▪ and ▪▪▪ indicate $p\leq0.01$ and 0.005 respectively). Bars in dark grey represent activity of the VNTR constructs in cortical cultures and light grey bars represent VNTR construct transcriptional activities in SN4741 cells.

Figure 4.9

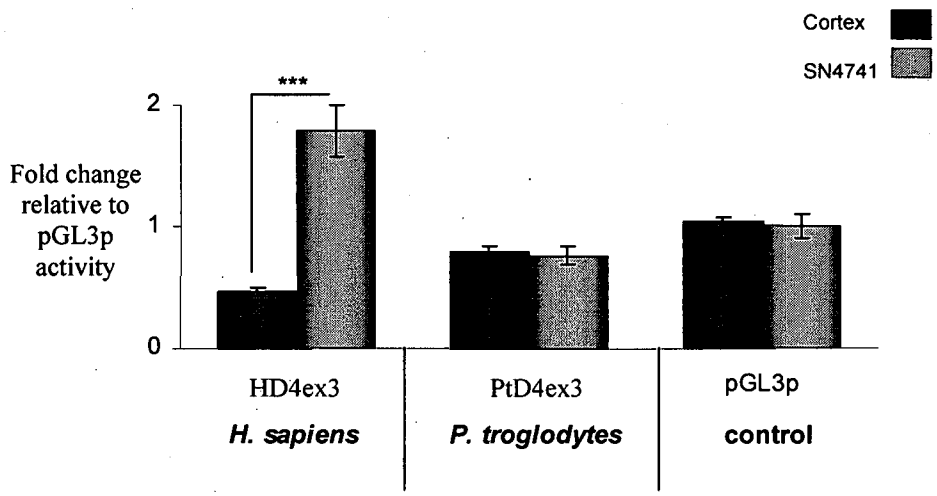


Figure 4.9 Comparison of the transcriptional abilities of the *H. sapiens* (HD4ex3) and *P. troglodytes* (PtD4ex3) VNTRs. The VNTR constructs were transfected into SN4741 cells (using TRANSFAST) and dissociated cultures of neonate rat cortex (using ExGen). Fold increase represents the basal level of expression of each constructs relative to pGL3p activity. The HD4ex3 construct supported significantly different levels of expression in SN4741 than in cortex ($p<0.005$). Values are given as mean and std. error based on 2 experiments (per triplicate, $n=6$).

Comparison of the levels of reporter gene expression supported by the *H. sapiens* and *P. troglodytes* DRD4 exon 3 VNTR constructs in the two *in vitro* models demonstrate that the activity of HD4.ex3 exhibits tissue specificity, which seems to be a common feature of the VNTRs tested in general (Table 4.8). In SN4741 cells the activity of the HD4ex3 construct was 3.91 fold more active than in dissociated frontal cortex cultures, and this difference in reporter gene expression levels was found to be very highly significant (Students *T*-test p=0.01, Table 4.8). Conversely, the marginal activity supported by PtD4ex3 constructs in SN4741 in both *in vitro* models did not vary (Table 4.8). This preliminary study suggests that the *P. troglodytes* D4ex3 VNTR construct may not harbour transcriptional regulatory activities in the cells analysed, and importantly, supports potential differences in the transcriptional activity exhibited by the *H. sapiens* D4ex3 VNTRs in different cells. However, since the activities of these regulatory elements are sensitive to cellular environment, testing of the D4ex3 VNTR and other polymorphism in the DRD4 gene of *H. sapiens* and *P. troglodytes* should be conducted.

Table 4.8

Average fold difference in luciferase activity in SN4741/cortex

construct name	HD4ex3	PtD4ex3	pGL3p
average fold increase over pGL3p	3.932 (***)	0.961	1.010
	±	±	±
standard error	0.469	0.084	0.028

Table 4.8. Ratio of reporter gene expression supported by HD4ex3 and PtD4ex3 in SN4741 cells/ activity supported in dissociated cultures of frontal. Fold increase represents the basal level of expression driven by each construct in SN4741 cells over the activity supported in dissociated cultures of neonate rat cortex (data taken from Table 4.7). Fold increase of expression levels exhibited by the *H. sapiens* HD4ex3 construct in the two models was highly significant (Student's *T*-test, ***=p=0.001).

4.4 Summary and brief discussion

This chapter aim to test 3 hypotheses which rise from the analysis of the VNTR sequences of chapter 3. These were:

(1) That the non-human primate VNTRs in the DRD4 (D4ex3) and in the SLC6A4 (STin2 and promoter) genes can act as *cis* regulators of gene expression, and could therefore contribute to the transcriptional regulation of these genes as proposed for the *H. sapiens* VNTRs.

(2) That the intra-specific variability in non-human and human primates SLC6A4 VNTR sequences (Figures 3.11 and 3.7) could contribute to the diversification of reporter gene expression, as a correlate for differential regulation of SLC6A4 expression within a population.

(3) That the great inter-specific variation observed between the sequences of *H. sapiens* and great apes VNTRs (Figures 3.10a and 3.13a) correlate with the variation in the transcriptional activities exhibited by the VNTRs of *H. sapiens* and other hominids in *vitro*.

In section 4.3.1, I demonstrated that intra-specific allelic variation observed in the VNTRs of SLC6A4 genes of *H. sapiens* and *P. troglodytes* correlates with diversification of reporter gene expression in dissociated cultures of neonate rat frontal cortex (Figures 4.1, 4.2 and 4.3). For example, the STin2 VNTRs of *H. sapiens* with either 9 and 12 repeat units supported significantly higher levels of reporter gene expression than the variant with 10 repeats. These findings are in agreement with previous observations of our group (Klenova et al., 2004; Roberts et al., 2007) and demonstrate that the observed differential regulation supported by these VNTR variants in clonal cell lines were also present in cultures derived from CNS. The activities of the *H. sapiens* STin2 VNTRs were higher in cortical cultures, and

this was also observed in the functional assays of the *P. troglodytes* STin2 VNTRs. Following from this analysis, I investigated the *cis* regulatory properties of the two most common alleles of the promoter VNTR found in the SLC6A4 gene of *H. sapiens* in dissociated cultures of neonate rat cortex. These VNTRs (Homo 16 and Homo 14) transcriptional activities were significantly different in the cortical cultures (Figure 4.3), and the ratio of their activities agrees with that observed in experiments conducted by others in different cell lines (Lesch et al., 1997; Sakai etal., 2002). However, in the present experiment these promoter VNTR variants acted as activators of reporter gene expression, being the short allele the highest expressor; contrastingly Lesch and Sakai observed that these VNTRs acted as repressors, being the long allele the highest expressor (Heils et al., 1997). This suggests that these VNTRs ability to support gene expression in neurons is distinct from their activities in other cell lines. In conclusion, these results demonstrate the capacity of these promoter and STin2 VNTRs in the SLC6A4 genes of *H. sapiens* and *P. troglodytes* to support tissue and sequence specific expression of a reporter gene in neuronal enriched cultures; these elements may therefore also act as *cis* regulators of the endogenous SLC6A4 genes of these species.

In section 4.3.2, I demonstrated that the variation of the SCL6A4 and DRD4 VNTR sequences of *H. sapiens* and great African apes observed in chapter 3 correlates with differences in the transcriptional activities their VNTRs exhibit in neuronal cultures. As an example, it was shown that during evolution, the sequences of the STin2 VNTR of *P. troglodytes* and *Gorilla* sp. had diverged greatly from the *H. sapiens* sequences (Figure 3.10a); in the functional analysis their VNTR constructs show a trend were the *H. sapiens* VNTRs support higher levels of expression than the *P. troglodytes* or *Gorilla* sp. (Figure 4.4). Conversely, the promoter VNTRs of

hominids had not diverged as much as the STin2 VNTRs, but importantly the sequences of *H. sapiens* and *P. troglodytes* were distinct. Similarly, the *in vitro* assay in cortical cultures demonstrated a significant distinction between the transcriptional activities of the 5' promoter VNTRs of *H. sapiens* and *P. troglodytes*. The functional analysis of the STin2 VNTR of cercopithecids (old world monkeys) suggests that the transcriptional regulation mediated by the STin2 VNTR originated around 35-25 million years ago, before hominids and cercopithecids separated. Indeed, the cercopithecid STin2 VNTRs supported robust reporter gene expression in neuronal cultures. Interestingly, comparison of the levels of transcriptional activities supported by the STin2 VNTR of *P. pygmaeus* and cercopithecids further suggests that the ancestral hominid STin2 VNTR exhibited levels of transcriptional activity similar to those of *P. pygmaeus*, *H. sapiens* and cercopithecids VNTRs (Figures 4.5 and 4.6).

The last section of this chapter investigated the *in vitro cis* regulatory properties of common variants of the VNTR in exon 3 of the DRD4 gene (D4ex3) of *H. sapiens* and *P. troglodytes*. Based on the differences in sequence identified between these species (Figure 3.21b), I hypothesised that the D4ex3 VNTRs of these species would exhibit differential *cis* regulatory activities. In section 4.3.2.6, I conducted a preliminary study, using one variant of the D4ex3 VNTR of *H. sapiens* and one of *P. troglodytes*. Functional analysis of these VNTR constructs demonstrated that only the *H. sapiens* variant acted as a tissue specific repressor/enhancer of reporter gene expression whilst PtD4.5 did not (Figure 4.9). It is possible that the same or other variants of *P. troglodytes* D4ex3 VNTR are active under other culture conditions, thus the regulatory properties of this VNTR in *P. troglodytes* cannot be discarded. However, it does demonstrate differences in their transcriptional activities between species. It is important to consider that the DRD4 gene presents abundant

polymorphisms, thus the analysis of the D4ex3 VNTRs using a heterologous system (luciferase reporter gene driven by a SV40 promoter) lacks other *cis* regulatory domains within the gene locus that may be necessary for the function of the *P. troglodytes* D4ex3 VNTR. In conclusion, this data is consistent with this polymorphic VNTR acting as a transcriptional regulator, contributing of the DRD4 gene expression, and to the diversification of the function of the DRD4 receptor proteins in dopaminergic neurons in the CNS of humans and non-human primates.

Neurotransmitter genes	**SLC6A4**		**DRD4**	**SLC6A3***	**Opioid * receptor**
Species	**STin2 VNTR**	**Promoter VNTR**	**D4ex3 VNTR**	**3'UTR VNTR**	**Promoter VNTR**
H. sapiens	high	high	high	low	high
P. troglodytes	low	low	low	high	low

Table 4.9 VNTRs of neurotransmitter genes of *P. troglodytes* and *H. sapiens* which have showing differential *cis* regulatory activity *in vitro*. VNTRs in neurotransmitter genes of *H. sapiens* and *P. troglodytes* tested in luciferase expression assays using heterologous and homologous systems show a trend of distinction between the transcriptional properties of these two species VNTRs. * Data taken from Rockman et al., 2005 and Inoue-Murayama et al., 2003.

In conclusion, the analysis of the transcriptional activities of these *cis* regulatory elements found the SLC6A4 gene and in the DRD4 gene of *H.sapiens* and *P. troglodytes* suggest that they exhibit different functional activities which could correlate with divergence of transcriptional regulation between these species. The evidence presented in this chapter is congruent with other studies conducted on VNTRs from other neurotransmitter genes correlated with cognitive abilities (Table 4.9). Put together, these findings are suggestive that as hypothesized, there is distinction between the activities of *cis* acting domains that mediate cognition and

emotional behaviour in *H. sapiens* and the great apes, particularly with *P. troglodytes.* This may in part explain the identified expression profiles differences between these two species (Khaitovich et al., 2004; Preuss et al., 2004) highlights the role of these highly plastic regulatory domains in the evolutionary adaptation the human brain underwent during the last 2 million years.

Chapter 5. Regulatory effects of CTCF and YB-1 on the primate VNTR.

5.1 Introduction

I*n vitro* experimentation conducted by our group has demonstrated that the STin2 VNTRs of *H. sapiens* are activated by TFs YB-1 and CTCF (Klenova et al., 2004, Roberts et al., 2007). This regulation is proposed to occur mainly via Y-binding sites for YB-1 or Y-boxes (Klenova et al., 2004; Kohno et al., 2003; Norman et al., 2001) and through sites for the binding of CTCF (Klenova et al., 2004; Roberts et al., 2007). Preliminary unpublished data from our group show that CTCF will also bind and modulate the function of the 5' promoter VNTR. Variation of these binding sites in the sequences of the *H. sapiens* STin2 VNTRs has been correlated with differential regulation mediated by YB-1 and CTCF (Kenova et al., 2004; Roberts et al., 2007). The demonstration of this regulation poses the question that, as the STin2 and 5' promoter VNTR sequences of human and non-human primate exhibit great sequence homology (as identified in chapter 3; Table 3.1, Figures 3.7 and 3.11) would CTCF and YB-1 mediate the regulatory function of the non-human primate VNTRs. Further, in chapter 1 (Figure 1.12) it was demonstrated that VNTRs in other neurotransmitter genes involved in primate behaviour shared homology with the STin2 VNTR, then is it possible that these other VNTRs are also regulated by CTCF and YB-1? Since these questions have not been investigated before, this chapter aimed to explore them by assessing the regulatory effects of CTCF and YB-1 on the transcriptional activities of different VNTRs found in the DRD4 and SLC6A3 gene *in vitro*.

5.2 Aims

The first aim of this chapter was to test the tissue specificity of the regulatory effects of CTCF and YB-1 on the transcriptional activities of the different variants of the *H. sapiens* STin2 VNTRs. For this I co-transfected the VNTR constructs and the expression vectors of YB-1 and CTCF into dissociated cultures of neonate rat frontal cortex and into a non-neuronal cell line (JAr cells) and compare the regulatory effects of CTCF and YB-1 on the reporter gene levels supported by the VNTR constructs.

The second aim of this chapter was to asses whether the intra and inter-specific variation identified in the primates STin2 and 5' promoter VNTR sequences correlated with differential regulation by YB-1 and CTCF *in vitro* in a neuronal environment. For this study, I co-transfected the STin2 and promoter VNTRs constructs (described in Table 4.1) and expression vectors of the human CTCF and YB-1 proteins into dissociated cultures of neonate rat cortex.

The third and last aim of this chapter was to investigate if, as predicted by sequence homology (Figure 1.12), YB-1 and CTCF can regulate the transcriptional activities of some other VNTRs in the SLC6A3 and DRD4 genes *in vitro*. This was achieved by co-transfection/nucleofection of CTCF and YB-1 expression vectors together with the different VNTR constructs and measurement of their reporter gene expression supported in neuronal and non-neuronal cell culture models.

5.3 Results

5.3.1 CTCF and YB-1 differentially regulate the transcriptional activities of the *H. sapiens* STin2 VNTRs in cultures of rat cortex and JAr cells.

Previous studies conducted by our group have suggested that the sequence variation of the STin2 VNTRs of *H. sapiens* may contribute to the differential regulation of the SLC6A4 gene expression in the CNS (e.g. Mackenzie and Quinn 1999; Fiskerstrand et al., 1999; Klenova et al., 2004; Roberts et al., 2007). This hypothesis is supported by the demonstration that CTCF and YB-1 differentially regulate the transcriptional activities of reporter gene constructs bearing the STin2 VNTRs in diverse cell lines of non-neuronal origin such as HEK293, COS7 and JAr cells (Klenova et al., 2004; Roberts et al., 2007). However, the potential differential regulation of the VNTR transcriptional activities has not been addressed in a cell culture model derived from the CNS. Therefore, I compared the regulatory effects of CTCF and YB-1 overexpression on the transcriptional activities of the *H. sapiens* STin2 VNTR reporter gene constructs in cultures of JAr and neonate rat cortex (Figure 5.1a and b).

The STin2 VNTRs reporter gene constructs and expression vectors of either CTCF or YB-1 were co-transfected as described in section 2.2.17. The ratio of CTCF or YB-1 expression vectors and the VNTR constructs was 1μg CTCF or YB-1 expression vectors per 1μg STin2 construct in both cell models (based on optimization studies not shown). The pGL3p empty vector was co-transfected with CTCF and YB-1 as a control and these factors had no effect on reporter gene expression (data shown in appendix 8). To compare the regulatory effects of CTCF and YB-1 on the VNTR constructs, I normalised the transcriptional activities of the VNTR constructs co-transfected with CTCF or YB-1 with the transcriptional

activities of the VNTR constructs transfected with equal amounts of pGL3b empty vector (section 2.2.17). The regulatory effects of YB-1 and CTCF were expressed as percentages of positive or negative regulation of reporter gene activity.

In JAr cells, over expression of CTCF induced up-regulation of the mean levels of reporter gene expression supported by both STin2.12 and STin2.9 constructs (Figure 5.1a), but this effect was only significant for the STin2.12 (Table 5.1). Conversely, the transcriptional activity of STin2.10 was not up-regulated by over-expression of CTCF, but instead was significantly repressed (Figure 5.1a, Table 5.1). Over-expression of YB-1 induced significant up-regulation of the STin2.9 activity (Figure 5.1a, Table 5.1) but did not affect the transcriptional activity of STin2.12 and also repressed the STin2.10 ability to support reporter gene expression (Figure 5.1a, Table 5.1).

In dissociated cultures of neonate rat cortex (Figure 5.1b), overexpression of CTCF and YB-1 induced repression of the basal transcriptional activities of all three STin2 VNTRs. However, the absolute amount of repression was variable. For example, CTCF induced greater repression of the transcriptional activities of STin2.9 and STin2.12 than of STin2.10 (Table 5.1). Similarly, YB-1 induced greater repression of the STin2.12 and STin2.9 transcriptional activities than the induced on the STin2.10 activities (Figure 5.1b, Table 5.1). The results suggest that there are differences in the regulation of YB-1 and CTCF on the variants of the STin2 VNTRs of *H. sapiens* in two models. However, both data sets consistently show that the STin2.10 regulation is distinct from that of STin2.9 and STin2.12, being less responsive to YB-1 and CTCF in cortical cultures and in JAr cells.

In conclusion, the results indicate that the STin2 VNTRs transcriptional activities are mediated by CTCF and YB-1 overexpression in a neuronal environment,

and that their effect is sequence and tissue specific, complementing previous findings conducted in other cell lines by our group (Klenova et al., 2004; Roberts et al., 2007).

Figure 5.1a

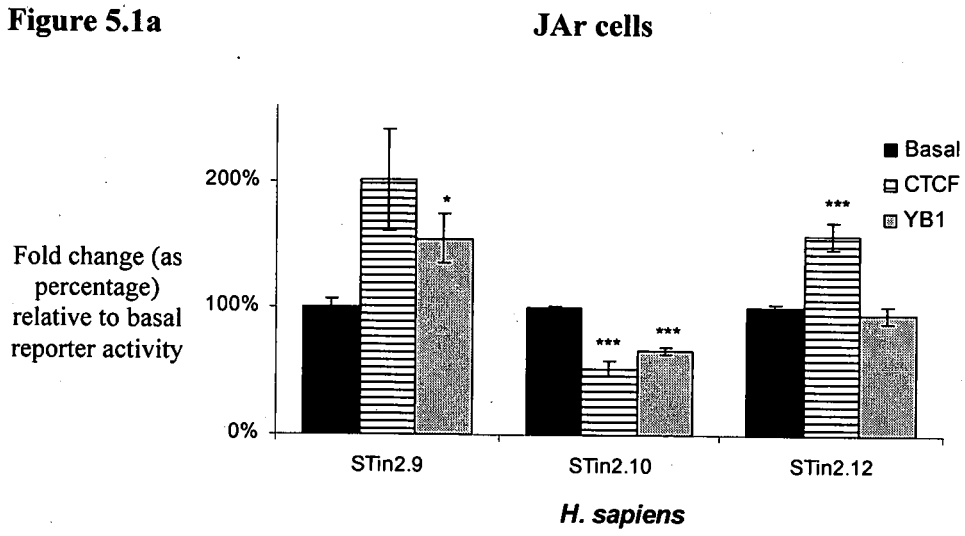


Figure 5.1b

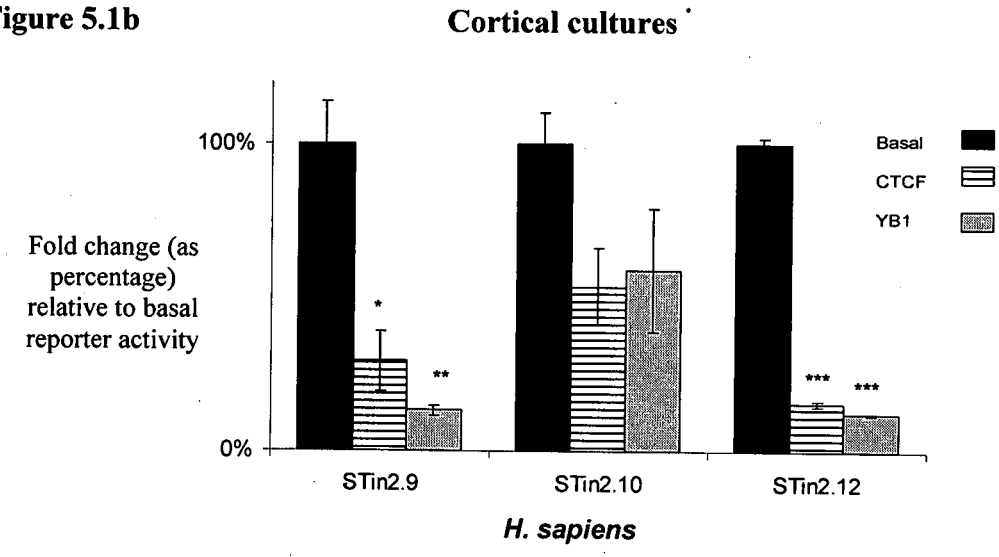


Figure 5.1 Regulatory effects of CTCF and YB-1 on the STin2 of *H. sapiens* in JAr cells and in cortical cultures. (a) The *H. sapiens* STin2.9, STin2.10 and STin2.12 plasmids were co-transfected with CTCF or YB-1 expression vectors (1μg of CTCF/YB-1 per 1 μg of VNTR construct) into JAr cells using TRANSFAST. Over expression of CTCF or YB-1 induced repression and activation of the basal transcriptional activities supported by the STin2 VNTRs. **(b)** The same constructs were co-transfected with CTCF or YB-1 expression vectors (1μg of CTCF/YB-1 per 1 μg of VNTR construct) into dissociated cultures of neonate rat cortex using ExGen 500. Co-transfection with YB-1 and CTCF induced repression of the transcriptional activities of the STin2 VNTR constructs. Repression or activation (expressed in percentages) was calculated as the fold change of each co-transfection relative to basal activity of the VNTR constructs. Co- transfections with YB-1 and CTCF were represented by dashed white bars grey bars respectively.

	Cortex			JAr		
construct name	STin2.9	STin2.10	STin2.12	STin2.9	STin2.10	STin2.12
CTCF	71*	47	85***	201	48 ***	158***
YB1	87**	42	88***	154*	33 ***	95

Table 5.1 CTCF and YB-1 differentially regulate the transcriptional activities of the *H. sapiens* STin2 VNTRs. CTCF and YB-1 overexpression differentially regulated the activities of the 3 STin2 VNTRs in *H. sapiens*. Values are given as mean based on 2 experiments (per triplicate). Repression or activation was calculated as the fold change (expressed as percentage) relative to basal reporter activity supported by the VNTR constructs. In the table *, ** and *** indicate significant differences between the levels of expression of each co-transfection with YB-1 and CTCF compared to the basal luciferase expression produced by the *H. sapiens* STin2 VNTR constructs (Student's *T*-test, * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.005$; $n=6$).

5.3.2 The African great ape STin2 VNTR constructs transcriptional activities are regulated by CTCF and YB-1?

The number of YB-1 and CTCF binding sites varies greatly between *H. sapiens* and the African apes STin2 VNTRs (Table 5.2). For example, the number of "g-Yboxes" (as defined in Klenova et al., 2004) varied from 7 in the *H. sapiens* VNTRs to 35 in that of *Gorilla* sp. (Table 5.2). In addition, the position and variation of "f" repeats which coincide with "a-Y boxes", were found to vary in different regions of the STin2 VNTRs of *H. sapiens* and *P. troglodytes* (Figure 3.7 and Table 3.1). Similarly, the number of CTCF binding sites (which coincide with "d" repeats) varies from 1 to 14 in the STin2 VNTRs of the latter species (Table 5.2).

Species name	number of a-Y boxes	number of repeats separating a-Y boxes	number of g-Y boxes	Number of CTCF binding sites per VNTR
Gorilla sp.	3	13 and 17 repeats	33 or 35	13 or 14
Pan troglodytes	1 or 2	8	12 or 13	1 to 6
Homo sapiens	1 or 2	4	7 or 8	3 or 4

Table 5.2. Differences in YB-1 and CTCF binding sites amongst *H. sapiens*, *P. troglodytes* and *Gorilla* sp. STin2 VNTRs. In these hominids STin2 VNTRs, the "g-Y boxes" (for YB-1 binding) were more closely distributed and more common than the a-Y boxes. The number of CTCF binding sites varied greatly, from 1 (in *P. troglodytes* VNTR) to 14 (in *Gorilla* sp.VNTR).

I therefore proposed that such sequence differences might correlate with potential variation in the regulation of hominids STin2 VNTRs by CTCF and YB-1. To test this hypothesis, I co-transfected YB-1 and CTCF expression vectors with the STin2 VNTR constructs of *P. troglodytes* and *Gorilla* sp. and compared to *H. sapiens* STin2 VNTR data (Figure 5.2). Protocols of co-transfections, cell culture preparations and measurement of transcriptional activities supported by the STin2 VNTRs were described before and hence will not be repeated further.

Results from the co-transfection experiments are shown in Figure 5.2 and Table 5.3. In dissociated cultures of neonate cortex, over expression of CTCF and YB-1 repressed the transcriptional activities of the STin2 VNTR constructs of *P. troglodytes* (STin2.18, STin2.19 and STin2.23) and *Gorilla* sp. (STin2.40) similar to that seen for the *H. sapiens* STin2 VNTRs. For example, CTCF overexpression induced a significant repression of the transcriptional activities of the STin2 VNTRs of *P. troglodytes*, that ranged from 40% to 82 %. Similarly, the activity of the *Gorilla* sp. VNTR construct was repressed in 54%. Over expression of YB-1 significantly repressed the transcriptional activities of the STin2 VNTRs of *P. troglodytes*, which ranged from 45% to 80%. The activity of the *Gorilla* sp. construct was repressed by YB-1 in 50% (Table 5.3).

In summary, the amount of repression induced by CTCF and YB-1 on the STin2 VNTR in *P. troglodytes* varied. Namely, STin2.19 transcriptional activity was significantly more repressed by either TF than STin2.18 or STin2.23. This result are consistent with the hypothesis that these VNTRs contribute to intra-specific variation of the regulation of SLC6A4 gene in *P. troglodytes* as proposed before, and this may be mediated by CTCF or YB-1 regulation. In addition, these results suggest that the STin2.40 of *Gorilla* sp., which did not support reporter gene expression (Figure 4.4b), may be functional as a repressor under different cell culture conditions. Finally, these results showed that the STin2 VNTRs of *P. troglodytes* and *Gorilla* sp. are down regulated by CTCF and YB-1 to similar levels as seen for *H. sapiens* VNTRs (Table 5.1), suggesting related mechanisms in their transcriptional regulation (mediated by YB-1 and CTCF) in this *in vitro* model.

Figure 5.2

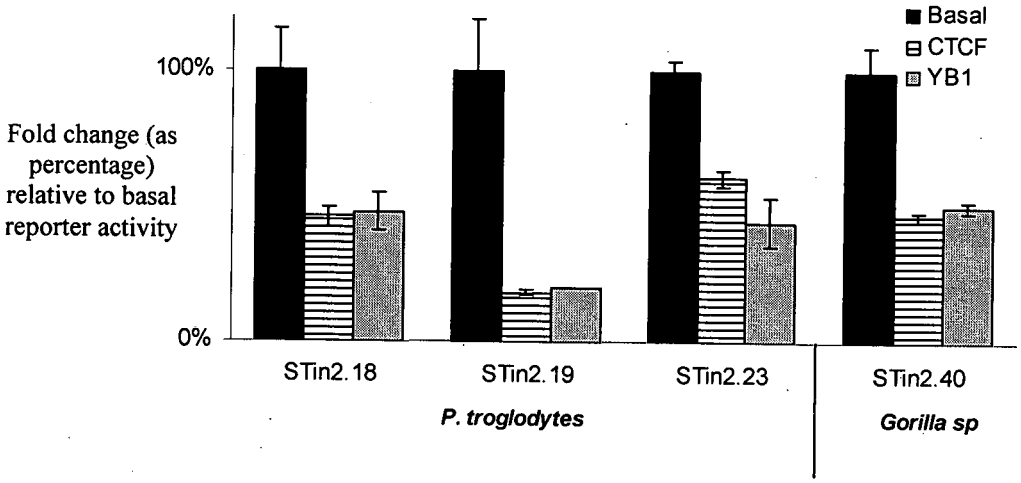


Table 5.3

Species	Construct name	CTCF repression	YB1 repression
P. troglodytes	STin2.18	54*	52*
	STin2.19	82*	80*
	STin2.23	40**	45*
Gorilla sp.	STin2.40	54**	50**

Figure 5.2 Regulatory effects of YB-1 and CTCF on the STin2 VNTRs of *P. troglodytes* and *Gorilla* sp. in cortical cultures. The STin2 VNTRs constructs were co-transfected with CTCF or YB-1 expression vectors (1μg of CTCF/YB-1 per 1 μg of VNTR construct) into dissociated cultures of neonate rat cortex using ExGen 500. Transcriptional activities of the VNTR constructs co-transfected with YB-1 are represented by dashed white bars and those co-transfected with CTCF are shown as grey bars

Table 5.3 CTCF and YB-1 significantly repressed the activities of the STin2 VNTRs of *P. troglodytes* and *Gorilla* sp. Values are given as mean based on 2 experiments (per triplicate). Repression (in percentage) was calculated as the fold change (expressed as percentage) relative to basal reporter activity suported by the VNTR constructs. In the table * and ** indicate significant differences between the levels of expression of each co-transfection compared to the basal luciferase expression produced by the VNTR construct (Student's *T*-test, * $p \leq 0.05$; ** $p \leq 0.01$; $n=6$).

5.3.3 CTCF and YB-1 repressed *P. pygmaeus* and cercopithecid STin2 VNTRs in dissociated cultures of neonate rat cortex.

Analysis of the sequences of the STin2 VNTRs found in *P. pygmaeus* and the cercopithecids *M .sphinx* and *C. aethiops* showed no evidence for " a-Y boxes", that were found in all other primate STin2 VNTRs analysed (Table 5.4). Similarly, the number of binding sites for CTCF also varied amongst these primates VNTRs, with the cercopithecids STin2 VNTR exhibiting 1 or 2 sites, and the *P. pygmaeus* STin2 VNTR exhibiting 3 sites. The variation in the number and position of "a-Y boxes" for the binding of YB-1 and the number of CTCF sites have been linked to differential regulation of the STin2 VNTRs of *H. sapiens* by YB-1 and CTCF (Klenova et al., 2004; Roberts et al., 2007). Therefore, these sequence differences could result in differential regulation of these and the other primate STin2 VNTRs by these factors.

Table 5.4

Species name	number of a-Y boxes	number of g-Y boxes	Number of CTCF binding sites
Pongo pygmaeus	0	6	3
Cercopithecus aethiops	0	5	1
Mandrillus sphinx	0	5	1

Table 5.4. Binding sites for YB-1 and CTCF in the STin2 VNTRs of *P. pymaeus*, *M. sphinx* and *C. aethiops*. In these VNTRs there is only "g-Y boxes" for the binding of YB-1, and the number varies from 5 to 6. The number of putative binding sites for CTCF varies from 1 to 3 in these species STin2 VNTRs.

To test whether the variation in TFBS for CTCF and YB-1 affected the regulation of the STin2 VNTRs of *C. aethiops*, *M. sphinx* and *P. pygmaeus*, I co-transfected the VNTR constructs with expression vectors of the human CTCF and YB-1 into dissociated cultures of neonate rat frontal cortex. The results are shown in Figure 5.3 and Table 5.5.

In brief, overexpression of both proteins repressed the transcriptional activities of the STin2 VNTRs (Figure 5.3). YB-1 induced repression of the transcriptional activities of the three VNTR constructs, and this repression was in the range of 63% to 86% for the cercopithecids VNTRs and of 55% on the *P. pygmaeus* VNTRs. CTCF overexpression also repressed the transcriptional activities of these constructs, with the activity of the *P. pygmaeus* VNTR construct being repressed in 71% in (Table 5.5), while the repression ranged from 69% to 82% for the cercopithecids VNTR.

In summary, the results indicate that levels of repression induced by YB-1 on the cercopithecids and hominids STin2 VNTRs were similar and ranged from 40% to 86%. Furthermore, the results suggest that the repression levels induced by YB-1 on the activities of all STin2 VNTRs tested were not correlated with the presence or position of "a-Y boxes" or the number of "g-Y boxes" found within the STin2 VNTR sequences in this cortical culture model. Similarly, the levels of repression induced by CTCF overexpression in the cercopithecids and *P. pygmaeus* STin2 VNTRs overlapped with the observed levels of repression induced by CTCF on the hominids STin2 VNTRs activities. This result suggest that in this model, the variation in the number of binding sites for CTCF found in the primate STin2 VNTR constructs tested did not correlate with the repression levels induced by the overexpression of this protein. However, as previously observed, the effect these sequence differences on the regulation by CTCF and YB-1 cannot be discarded, since these regulation can change under other cellular conditions, e.g. stress, specific stimuli.

Figure 5.3

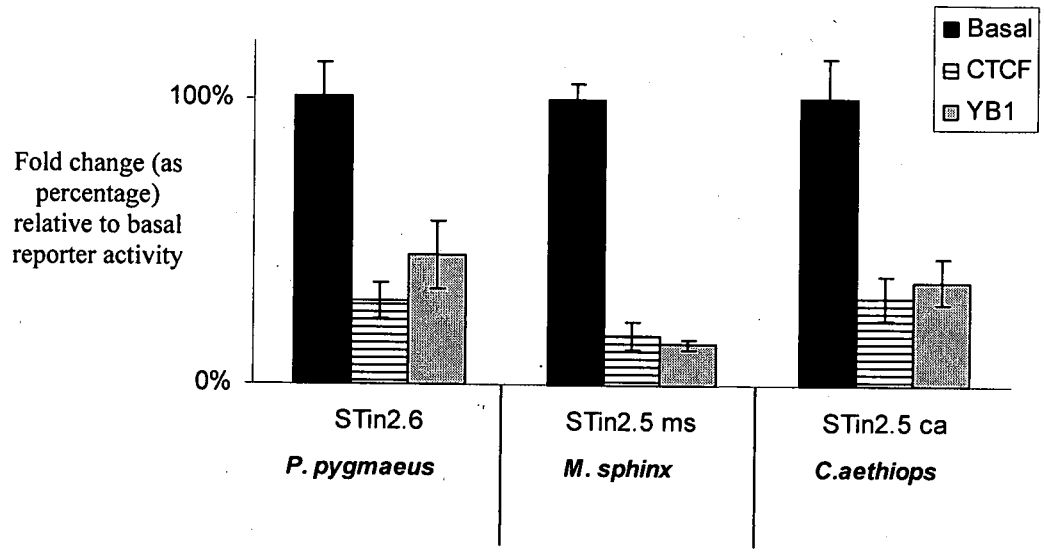


Table 5.5

Species	Construct name	CTCF repression (in percentage)	YB1 repression (in percentage)
P. pygmaeus	STin2.6	71***	55**
M. sphinx	STin2.5ms	82***	86***
C. aethiops	STin2.5ca	69***	63***

Figure 5.3 Regulatory effects of YB-1 and CTCF on the STin2 VNTRs of *P. pymaeus*, *M. sphinx* and *C. aethiops* in cortical cultures. Plasmids were co-transfected with CTCF or YB-1 expression vectors (1μg of CTCF/YB-1 per 1 μg of VNTR construct). Transcriptional activities of the VNTR constructs co-transfected with YB-1 are represented by dashed white bars and those co-transfected with CTCF are shown as grey bars.

Table 5.5 Repression induced by CTCF and YB-1 on the STin2 VNTRs. Values are given as mean based on 3 experiments (per triplicate). Repression (in percentage) was calculated as the fold change (expressed as percentage) relative to basal reporter activity suported by the VNTR constructs. In the table ** and *** represents significant differences between the levels of expression of each co-transfection compared to the basal luciferase expression produced by the VNTR construct (Student's *T*-test, * $p \leq 0.05$; ** $p \leq 0.01$).

5.3.4 CTCF and YB-1 regulate the promoter 5'VNTR of the SLC6A4 gene of primates.

Regulatory domains mediating a response to the same signalling pathway are likely to be regulated in part by the same TFs. In such cases, distinct functional VNTRs within one gene such as the SLC6A4 promoter and STin2 VNTRs may act synergistically to modify gene expression. To investigate this possibility, I co-transfected VNTR constructs bearing modern human and non-human primate 5' promoter VNTRs (described in Table 4.1) with expression vectors expressing CTCF and YB-1 *in vitro* in dissociated cultures of neonate rat frontal cortex.

5.3.4.1 *H. sapiens* SCL6A4 promoter VNTRs are differentially regulated by CTCF and YB-1 in rat neonate frontal cortex

The results of the co-transfection experiments are shown in Figure 5.4 and Table 5.6. In brief, both CTCF and YB-1 repress the transcriptional activities of the *H. sapiens* 5' promoter VNTR constructs. For example, CTCF overexpression repressed the transcriptional activities of the Homo 14 and Homo 16 in 71% and 38% respectively. This repression was only statistically significant for the activities of Homo 14 (Table 5.6). YB-1 overexpression repressed the average transcriptional activities of both constructs (by 63% and 50% of Homo 14 and Homo 16 respectively, Table 5.6); however, these effects were not statistically significant. Although the amount of repression induced by either TF on the Homo 14 and Homo 16 appear different, the repression levels are very similar. Thus, it is possible that the variation of the experimental conditions, which is characteristic of primary cultures could affect the significance of these results.

Figure 5.4

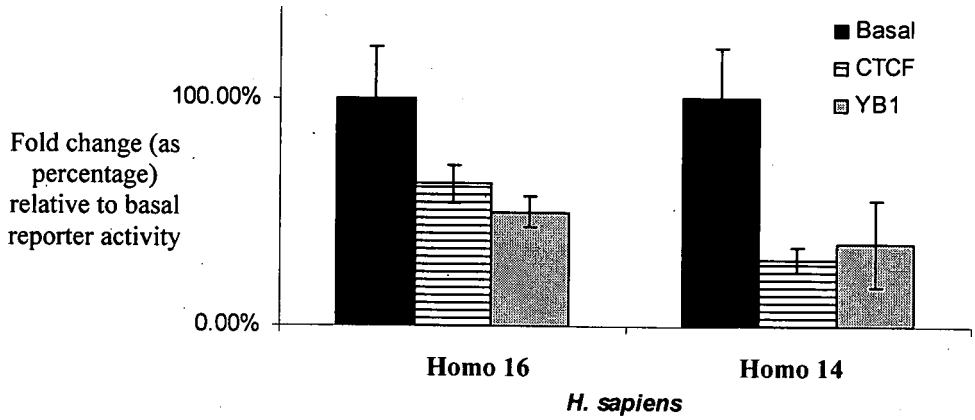


Table 5.6

Species	Construct name	CTCF repression (in percentage)	YB1 repression (in percentage)
H. sapiens	Homo 16	38	50
	Homo 14	71*	63

Figure 5.4 Regulatory effects of YB-1 and CTCF on the activities of the promoter VNTRs of *H. sapiens* in cortical cultures. Plasmids were co-transfected under basal conditions with CTCF or YB-1 expression vectors (1μg of CTCF/YB-1 per 1 μg of VNTR construct). Repression induced by CTCF and YB-1 are given as mean percentage of repression based on 3 experiments (per triplicate).

Table 5.6 CTCF and YB-1 repress the activities of Homo 16 and Homo 14 constructs. The level of expression of each co-transfection is calculated as fold change (expressed as percentage) relative to basal reporter activity suported by the VNTR constructs. In the table * indicate significant differences between basal and constructs co-transfected with CTCF (Student's *T*-test, * $p \leq 0.05$).

5.3.4.2 CTCF and YB-1 regulate the non-human primate promoter VNTR in dissociated cultures of rat neonate frontal cortex

The analysis of the SLC6A4 5'promoter VNTRs of hominids demonstrated that their sequences are highly homologous (Figure 3.11). When transfected into cortical cultures, the basal transcriptional activities of the transfected hominid promoter VNTR constructs showed an overlap (Figure 4.8). The exception was the *P. troglodytes* construct (Pt17.5) which unlike other constructs, did not exhibit transcriptional activity. The homology of sequences suggest that the function of the non-human primate 5' promoter VNTRs may be mediated by CTCF and YB-1, as observed for the *H. sapiens* VNTRs (Figure 5.4 and unpublished data Ali et al., manuscript in preparation). Furthermore, the sequence variation suggests potential variation in the regulation effected by these TFs. I investigated this question by co-transfecting the promoter VNTR constructs of *P. troglodytes*, *Gorilla* sp. and *P. pygmaeus* (described in Table 4.1) into dissociated cultures of neonate rat cortex and comparing their activities *in vitro*.

The results are presented in Figure 5.5 and Table 5.7. Briefly, YB-1 and CTCF differentially repressed the function of the 5' promoter VNTR constructs. For example, over-expression of CTCF induced repression of the basal activities of *Gorilla* sp, *P. pygmaeus* and *P. troglodytes* constructs, but this effect was not significant for the *P. troglodytes* construct (Table 5.7). YB-1 overexpression induced significant levels of repression to the 3 VNTRs constructs tested (Table 5.7). In summary, the results suggest that in general, the activities of the promoter VNTRs of hominids are responsive to regulation by YB-1 and CTCF in this cell model, and this may reflect regulation of these VNTRs *in vivo*.

It is noteworthy that the overexpression of YB-1 and CTCF repressed the 5' promoter VNTR construct of *P. troglodytes,* which did no exhibit transcriptional activity under basal conditions (Figure 4.8b). This suggests that this VNTR variant may be only active when the stress-signalling pathway of neurons is activated in the cell, in which YB-1 is involved (Klenova et al., 2004; Kohno et al., 2003; Norman et al., 2001). In conclusion, the levels of repression induced by YB-1 and CTCF on the 5' promoter VNTRs tested (of human and non-human primate origin) were similar.

Figure 5.5

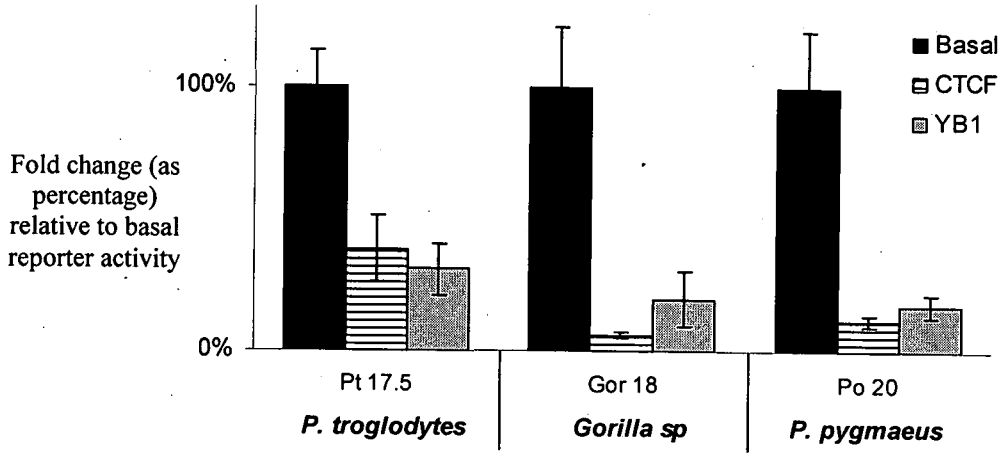


Table 5.7

Species	Construct name	CTCF repression (in percentage)	YB1 repression (in percentage)
P. troglodytes	Pt 17.5	62	70**
Gorilla sp.	Gor 18	94*	80*
P. pygmaeus	Po 20	89*	83**

Figure 5.5 Regulatory effects of YB-1 and CTCF on the activity of the promoter VNTRs of *P. troglodytes* and *Gorilla* sp. in cortical cultures. Plasmids were co-transfected with CTCF or YB-1 expression vectors (1μg of CTCF/YB-1 per 1 μg of VNTR construct). Both TFs repressed the basa transcriptional activities of the VNTR constructs.

Table 5.7 CTCF and YB-1 induced repression of the function of the hominids promoter VNTRs. The effect of CTCF and YB-1 is expressed as the average Fold change (expressed as a percentage) relative to basal reporter activity. Average values are based on 3 experiments (per triplicate) (Student's *T*-test, * $p \leq 0.05$; ** $p \leq 0.01$; $n=9$).

5.3.5 CTCF and YB-1 regulates the transcriptional activities of VNTRs in neurotransmitter genes with related sequences to the STin2 VNTR

In chapter 1, I constructed an alignment of the STin2 VNTRs and others located in the DRD4 and SLC6A3 genes, which have been linked to variation of modern human behaviour (Figure 1.12). The alignment showed that all these sequences shared homology, which is suggestive of commonalities in the regulation of their function. In a previous publication, portions of the sequence of the STin2 and the 3'UTR VNTRs of the SLC6A4 and SLC6A3 genes respectively were shown to be able to bind to similar protein complexes (Michelhaugh et al., 2001). This evidence suggests that the regulation of these VNTRs in different neurotransmitter genes may be mediated by CTCF and YB-1. To investigate this question, I analysed the potential regulatory effects of CTCF and YB-1 on the function of the DRD4 and SLC6A3 VNTRs in dissociated cultures of neonate rat cortex (DRD4 VNTRs) and midbrain (SLC6A3 VNTR).

5.3.5.1 Transcriptional activities of the SLC6A3 3'UTR VNTRs are repressed by overexpression of CTCF and YB-1

I analysed the regulatory effects of CTCF and YB-1 overexpression on the transcriptional activities of the two most common copy number variants (with 9 and 10 repeat units) of the 3'UTR found in the *H. sapiens* SLC6A3 gene (DAT 9 and DAT 10) when transfected into dissociated cultures of neonate rat midbrain. This region was selected due to the known expression of the SLC6A3 gene in the ventral tegmental area (VTA) and substantia nigra (SN) located in the midbrain (Shimada et al., 1992). For these experiments, I used nucleofection (using the AMAXA machine) instead of ExGen 500® as the DNA delivery method. Due to restrictions in the volume to be transfected using this method, the ratio of expression vectors expression

CTCF and YB-1 vs. VNTR constructs was changed to 0.5μg of TF: 1μg of VNTR construct. The transcriptional activities of the VNTR constructs were assessed as described before (Figures 2.2.16.2 and 2.2.17).

The basal levels of transcriptional activities of these VNTRs have not been studied in neuronal enriched cultures, thus these were shown in Figure 5.6a. In summary, in midbrain cultures the DAT 9 significantly repressed the expression supported by pGL3p alone, whereas DAT 10 did not (Students' *T*-test pGL3p vs. DAT9 p=0.03, vs. DAT10 p=0.1). Over-expression of the human CTCF and YB-1 induced strong levels of repression of the basal activity supported by both DAT 9 and DAT 10 constructs (Figure 5.6b, Table 5.8). There was no significance difference in the repression induced by CTCF and YB-1. CTCF induced significant repression of similar levels on the DAT 9 and DAT 10 (93% and 100% respectively), and YB-1 induced 90 and 98% of repression of the expression suported by DAT 9 and DAT 10 (Table 5.8).

The results suggest that the CTCF and YB-1 could act as repressors of these VNTRs in the CNS, as predicted by their sequence homology with STin2 VNTR. The differences in the sequences of these two variants of the DAT 3'UTR VNTR were not sufficient to invoke differential regulation by these TFs. However, other experimental methods to answer this question should be used in future experiments.

Figure 5.6a

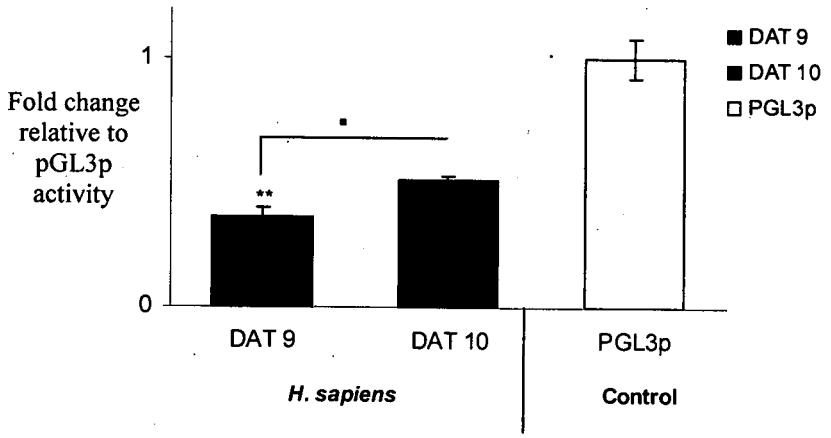


Figure 5.6b

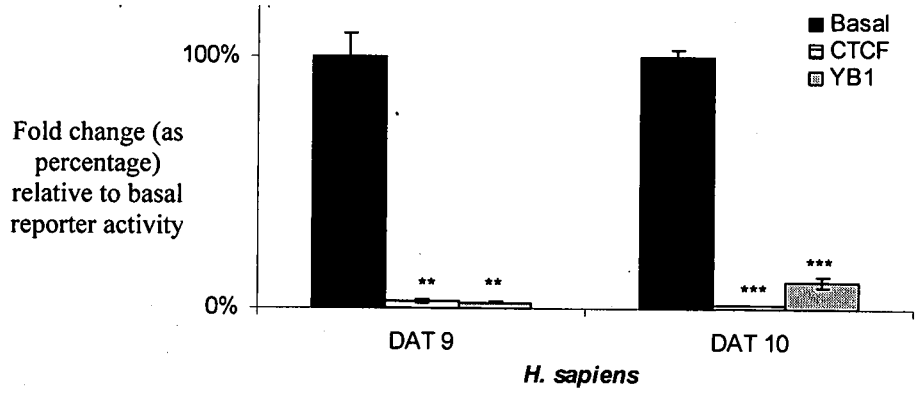


Figure 5.6a Transcriptional activities of the SLC6A3 3'UTR VNTRs (DAT 9 and 10) in dissociated cultures of rat midbrain. (a) The two constructs were nucleofected into midbrain cultures using the Nucleofector device (AMAXA). Only DAT 9 exhibit basal activity (as a repressor) when compared to the control (pGL3p) (See Table 5.8).

Figure 5.6b Regulatory effects of YB-1 on the activities of the DAT 9 and DAT 10 constructs. Constructs were co-nucleofected under basal conditions with CTCF or YB-1 expression vectors (1μg of CTCF/YB-1: 1μg of VNTR construct) using the Nucleofector device (AMAXA). Both TFs significantly repressed the activities of the DAT 9 and DAT 10 constructs (see Table 5.8).

Table 5.8

Species	Construct name	Basal levels	CTCF repression (in percentage)	YB1 repression (in percentage)
H. sapiens	**DAT 9**	0.359 ± 0.032 (•**)	97**	98**
	DAT 10	0.508 ± 0.015	99***	90***

Table 5.8 DAT 9 and DAT 10 transcriptional activities are repressed by CTCF and YB-1 overexpression. Basal transcriptional activities supported by DAT 9 and DAT 10 constructs were calculated as the the fold change relative to pGL3p activity. Their activities were significantly different from pGL3p (where • indicates by p=0.05) Values given are means and std. error based on 3 experiments (per triplicate). CTCF and YB-1 significantly repressed the basal transcriptional activities of DAT9 and DAT 10. In the table ** and *** indicate significant differences between constructs co-transfected with either CTCF or YB-1 compared to their basal activities (Students *T*-test, **=p≤0.01; ***= p≤0.05, *n*=6).

5.3.5.2 Regulatory effects of CTCF and YB-1 overexpression on the basal transcriptional activities of D4 ex3 VNTRs of *H. sapiens* and *P. troglodytes*

In chapter 4, I demonstrated that one of the variants of *H. sapiens* and *P. troglodytes* (HD4ex3 and PtD4ex3 respectively) supported differential reporter gene expression in dissociated cultures of neonate rat cortex (described in section 4.3.2.6). These DRD4 VNTRs share homology with the STin2 VNTRs of the SLC6A4 gene (Figure 1.12) suggesting that CTCF and YB-1 could be regulators of this VNTRfunction in neuronal cultures, as seen previously for the SLC6A3 VNTRs (section 5.3.5.1). However, the differences in the primary sequences of the HD4ex3 and PtD4ex3 could influence the interaction between the TFs. To investigate this hypothesis, I conducted a co-transfection assay of expression vectors expressing CTCF or YB-1 and the HD4ex3 and PtD4ex3 constructs as described in section 2.2.17.

The results of the co-transfection experiments are shown in Figure 5.7. Briefly, over expression of CTCF and YB-1 produced different levels of repression of the transcriptional activities of both VNTR constructs (HD4ex3 and PtD4ex3, Figure 5.7). For example, the repression induced by CTCF on both constructs was similar and statistically significant (Table 5.9). Conversely, overexpression of YB-1 only affected the construct bearing the PtD4ex3 VNTR (inducing repression of 50%), whereas the *H. sapiens* construct was not affected (Figure 5.7, Table 5.9). This difference was statistically significant. These results indicate the sequence variation that exist amongst the sequences of these two VNTR constructs (appendix 9) are capable of invoking differential regulation by CTCF and YB-1. Interestingly, CTCF regulated the D4ex3 VNTR of *H. sapiens*, as predicted by sequence homology (Figure 1.12) but the variation between the sequences of this VNTR and the STin2 VNTRs resulted in variation in the regulation by YB-1.

Figure 5.7

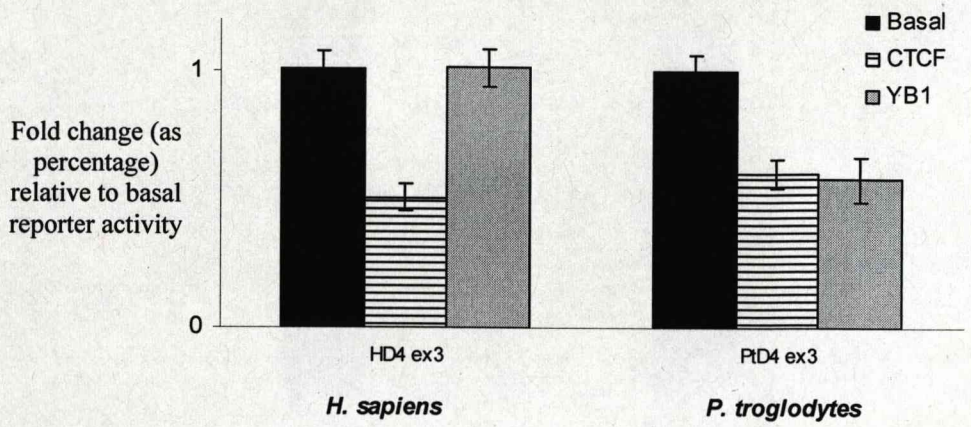


Table 5.9

Species	Construct name	CTCF repression (in percentage)	YB1 repression (in percentage)
P. troglodytes	PtD4 ex3	40***	42***
H. sapiens	HD4 ex3	50***	0

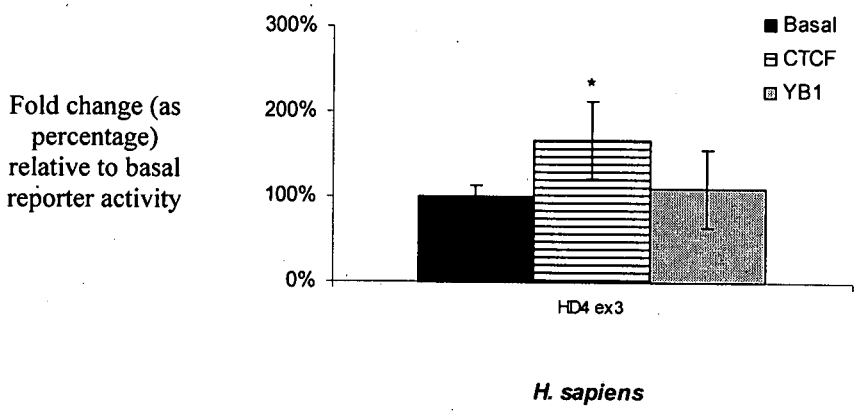
Figure 5.7 Regulatory effect of CTCF and YB-1 on the activities of D4ex3 VNTR constructs of *H. sapiens* and *P. troglodytes* in cortical cultures. The plasmids were co-transfected with CTCF or YB-1 expression vectors (1μg of CTCF/YB-1 per 1 μg of VNTR construct) using ExGen 500. CTCF repressed the transcriptional activities of HD4ex3 and PtD4ex3 constructs, but YB-1 only repressed PtD4ex3.

Table 5.9 Differential repression of YB-1 on the HD4ex3 and PtD4ex3 VNTR constructs. The repression induced by CTCF and YB-1 are calculated by normalising the luciferase values of each co-transfection with the basal luciferase expression supported by either VNTR constructs. In the table *** indicate significant differences between basal transcriptional activities and co-transfections with either CTCF or YB-1 (Student's *T*-test, *** $p \leq 0.005$, $n=9$).

I further explored the differential regulation of the HD4ex3 VNTR construct by YB-1 and CTCF in a non-neuronal cell line model (JAr cells, Figure 5.8). In brief, overexpression of the CTCF and YB-1 proteins in JAr cells induced a different effect (to the observed in cortical cultures) on the transcriptional activity of the HD4ex3 VNTR construct. CTCF induced an up-regulation of the basal transcriptional levels of the HD4ex3 of 160% whilst overexpression of YB-1 had not effect on the transcriptional activities of this construct (Figure 5.8).

In conclusion, these experiments showed that CTCF and not YB-1 acted as a regulator of this variant of the D4ex3 VNTR of *H. sapiens* under the two different culture conditions. Finally, the repression of YB-1 on the PtD4ex3 construct in the cortical cultures also suggest differential regulation mediated by TFs occur between these VNTR constructs from *H. sapiens* and *P. troglodytes*.

Figure 5.8



5.8 Regulatory effects of CTCF and YB-1 on the activities of the HD4ex3 VNTR construct on JAr cells. The HD4ex3 construct was co-transfected with CTCF or YB-1 expression vectors (1μg of CTCF/YB-1 per 1 μg of VNTR construct) using TRANSFAST. Regulatory effects (express in percentage) were calculated as the fold change (as expressed as a percentage) relative to basal reporter activity of the VNTR construct alone. In the graph, * indicates significant differences between basal and constructs co-transfected with CTCF (T-test, * $p \leq 0.05$, $n=9$).

5.3.5.3 Transcriptional activities of *H. sapiens* promoter VNTR in the DRD4 gene are repressed by over-expression of CTCF

Seaman (1999) identified a VNTR located in the 5' promoter region of the DRD4 gene of primates (with 1 or 2 repeat units per VNTR, 120 bp/repeat units). The variants of this VNTR have been shown to support differential reporter gene expression *in vitro* (D'Souza et al., 2004), and association studies have linked variation of this VNTR locus to differential regulation of the DRD4 expression in *H. sapiens* (e.g. MacCraken et al., 2000). Examination of the primary sequences of the two variants show no evidence for binding sites for either YB-1 or CTCF. However, I identified sites for Sp1 (appendix 10). As CTCF has been shown to be able to bind Sp1 sites in the c-myc promoter of chicken (Lobanenkov et al., 1990), and as CTCF multiple zinc fingers allows it to bind to several consesus sequences (Ohlsson et al., 2001), it is possible that CTCF binds and regulates the D4promoter VNTR. Furthermore, in the previous section it was shown that the CTCF protein might mediate the function the HD4ex3 VNTR also located in the DRD4 gene, the promoter VNTR constructs (D4L and D4S) may be regulated by this protein as well. I tested this hypothesis by co-transfecting constructs bearing the two variants of the 5' promoter VNTR of the modern human DRD4 gene (D4l and D4S) with the expression vector of human CTCF in dissociated cultures of neonate rat cortex as previously described.

First, it was established whether the DRD4 promoter constructs were capable of supporting reporter gene expression under basal culture conditions (Figure 5.9). In brief, the construct were not active in dissociated cultures of neonate rat cortex. This result is consistent with data from other VNTRs tested in this thesis, which showed that some VNTRs have restricted transcriptional activity in neuronal cultures (e.g.

STin2.40 in Figure 4.8b). However, the results of the co-transfection experiments with CTCF (Figure 5.10) indicate that under the influence of specific stimuli, these VNTR constructs exhibit transcriptional activity as repressors. The transcriptional activities of the D4 promoter constructs were significantly repressed by CTCF (Student's *T*-test, p=0.05 Table 5.10), as observed in the HD4ex3 VNTR (Figure 5.7). However, the repression induced by CTCF on the activities of both constructs was similar (79% and 86% respectively). This study suggests that CTCF may mediate the transcriptional activities of the DRD4 5' promoter VNTRs in neuronal cultures as deduced by the presence of Sp1/CTCF sites in the VNTR sequences. The down regulation induced by CTCF offers further support for the hypothesis proposed, that these elements might only be active under specific cellular conditions such as cellular stress, in which CTCF has been demonstrated to participate (Ohlsson et al., 2001; Wu et al., 2006). To corroborate this hypothesis, future experiments of the interactions of CTCF and these VNTRs *in vitro* and *in vivo* should be conducted.

Figure 5.9

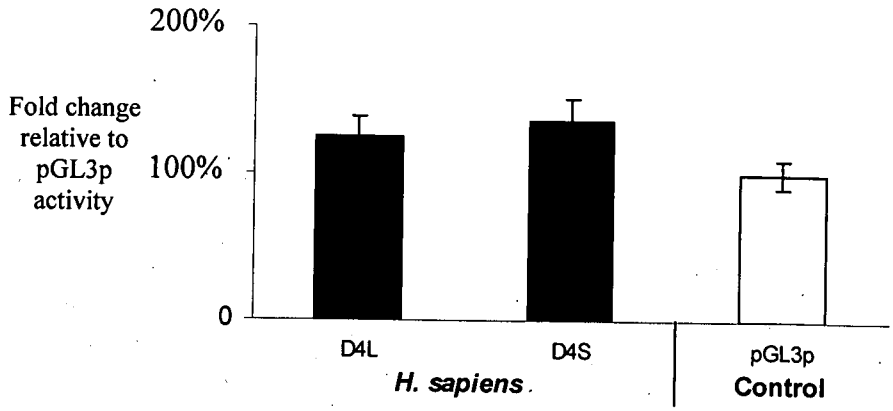


Figure 5.9 Basal transcriptional activities of the DRD4 promoter VNTRs of *H. sapiens* in cortical cultures. VNTR constructs were transfected under basal conditions using ExGen 500. Basal transcriptional levels of the D4L and D4S constructs were calculated as the transcriptional activities supported by each VNTR construct over pGL3p activity. Neither construct significantly affected the basal activity of the control (pGL3p) (See Table 5.10).

Figure 5.10

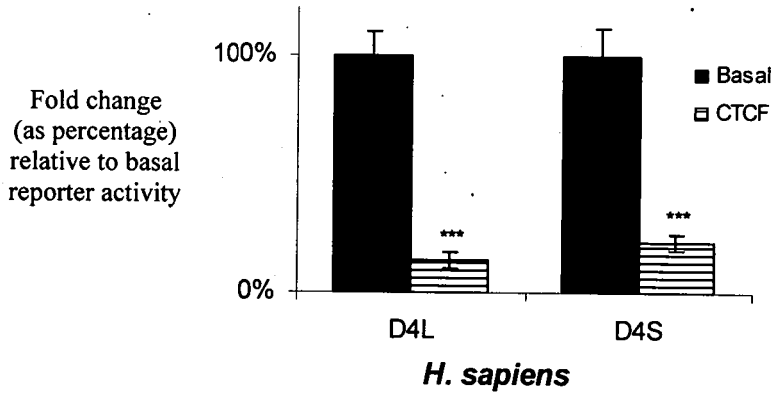


Table 5.10

	D4L		D4S		Control
	Basal fold increase (pGL3p)	repression induced by CTCF in %	Basal fold increase (pGL3p)	repression induced by CTCF in %	pGL3p
average value	1.253	86***	1.352	79***	1.000
	±	±	±	±	±
SE	0.125	3.3	0.160	3.1	0.064

Figure 5.10 Regulatory effects of CTCF on the activities of the D4L and D4S construct on cortical cultures. VNTR constructs were co-transfected under basal conditions with the CTCF expression vector (1μg of CTCF per 1μg of VNTR construct) using ExGen 500. CTCF significantly repressed the transcriptional activity of both VNTR constructs (Student's *T*-test, *** $p \leq 0.005$, $n=9$).

Table 5.10 D4L and D4S transcriptional activities are repressed by CTCF overexpression. Regulatory effects of CTCF were calculated as the fold change (expressed as a percentage) relative to reporter activity of the VNTR construct basal activities. In the table *** indicate significant differences between the basal activities of the VNTR constructs and their activities when co-transfected with CTCF (Student's *T*-test, *** $p \leq 0.005$, $n=9$). Basal and co-transfection values are means based on 3 experiments (per triplicate).

5.4 Summary and brief discussion

This chapter aimed to address three questions concerning the regulation of the function of selected VNTRs found in neurotransmitter genes by the TFs CTCF and YB-1. The first hypothesis tested was that in neuronal derived cell environment, both CTCF and YB-1 could regulate the STin2 VNTRs of *H. sapiens* and, the nature of the regulatory effects are tissue specific. Section 5.3.1.1 demonstrated that indeed, CTCF and YB-1 overexpression significantly regulated the the transcriptional activities of the STin2 VNTRs of *H. sapiens* in a neuronal environment, resulting in repression of their activities (Figure 5.1b). These effects where indeed tissue specific, as when transfected into JAr cells, the regulatory effect was reversed to enhancement (Figure 5.1a). The results also complement previous findings of our group, showing that the regulation of CTCF and YB-1 on this VNTR is sequence dependent (Lovejoy et al., 2003). For example, studies in both cell models consistently showed that the function of the STin2.10 construct was less affected CTCF and YB-1 (Table 5.1) than the other reporter gene constructs. These differences in regulation may be correlated to the variation in number and distribution in putative Y-boxes for YB-1 binding and binding sites for CTCF that exist amongst these three STin2 VNTR variants (appendix 11).

The results of co-transfections in JAr cells presented in this thesis differ from previous publications generated by our group (Roberts et al., 2007). It is possible that the difference in the amount of TF expression vectors used here and in previous publications (1:1.5μg respectively) could have affected the ratio of YB-1:CTCF in a cell, resulting in variation of activation or repression as previously observed in a variety of cells (JAr, HEK293, COS7 Klenova et al., 2004; Roberts et al., 2004). Nevertheless, the presented findings suggest important differences in the regulation of

the STin2 VNTRs mediated by CTCF and YB-1 in a neuronal and a non-neuronal cell models both expressing endogenous SLC6A4 gene (Roberts et al., 2007 and appendix 2). This suggest that under basal conditions, this VNTR could act as enhancers of gene expression in the CNS, but under stress (mediated by CTCF and YB-1), the activities of these may be repressed affecting the levels of expression of the SLC6A4 gene. However, it is important that other regulatory domains found in this gene locus such as the 5'promoter VNTR be considered, as these are likely to act in synergy with the STin2 VNTRs to regulate the expression of the SLC6A4 gene in the CNS.

The second aim of this chapter was to test whether the intra-specific and inter-specific variation of the sequences of the primate STin2 VNTR and 5' promoter could affect their regulation by CTCF and YB-1 in cell cultures derived from the CNS. The results of sections 5.3.2 and 5.3.3 demonstrate that the transcriptional activities of all STin2 VNTR constructs of hominids and cercopithecids tested are repressed by CTCF or YB-1 (Figures 5.2 and 5.3). The experiments also indicate that the variation in CTCF and YB-1 sites amongst the *P. troglodytes* STin2 VNTR constructs tested affect the interactions with CTCF and YB-1 (Figure 5.2 and Table 5.3) as the STin2.19 construct is more down-regulated by both YB-1 and CTCF than the other 2 VNTR reporter gene constructs. Conversely, the inter-specific differences affecting number, distribution and type of binding sites for YB-1 and CTCF found amongst the hominids STin2 VNTRs (described in Tables 5.2 and 5.4), did not correlate with differential regulation of the VNTRs by either TF. This lack of correlation could have been caused by several reasons, one of them being that the TFs induced repression on constructs with low levels of basal transcriptional activities (e.g *Gorilla* sp. STin2.40 in Figure 4.4b). Thus, the sensitivity of the luciferase assays may not be sufficient to

identify potential differential repression induced by the either CTCF or YB-1 in this cell culture model.

I analysed if some variants of modern human and non-human primate promoter VNTRs of the SLC6A4 gene were regulated by CTCF or YB-1, known to regulate the function promoter VNTRs in a non-neuronal cell line (Ali et al., manuscript in preparation). The results of section 5.3.4.1 and 5.3.4.2 showed that overexpression of either CTCF or YB-1 repressed the transcriptional activities of all 5' promoter VNTRs tested. These results are noteworthy, as it was proposed that CTCF and YB-1 would mediate the function of VNTRs such as STin2 and promoter VNTRs involved in the regulation of the same gene. These experiments did not demonstrate that the differences in putative CTCF and YB-1 sites that, exist in the *H. sapiens* 5' promoter VNTRs mediate differential regulation via CTCF and YB-1 in cultures derived from the CNS, neither had these experiments demonstrated that as predicted by their sequence variation, these TFs differentially regulated the 5' promoter VNTRs of hominids. However, CTCF and YB-1 overexpression did repress VNTR constructs that under basal culture conditions exhibited low or no activity (Figure 4.8b), such as the *P. troglodytes* or *P. pygmaeus* 5' promoter VNTR constructs (Figure 5.5). Thus, it is possible that as seen for the STin2 VNTRs, under cellular stress or specific challenge, these inactive VNTRs could be differentially regulated.

The final hypothesis tested in this chapter was if YB-1 and CTCF regulate the functions of other VNTRs characterised by GC rich sequences that are linked to the diversification of primate behaviour and shared homology with the STin2 VNTR sequence (D4ex3 and promoter VNTRs of the DRD4 gene and 3' UTR VNTR of the SLC6A3 genes; Figure 1.12). I tested this hypothesis by co-transfecting some

common variants of these VNTRs with expression vectors expression human CTCF and YB-1. Analysis of the regulatory effects of YB-1 and CTCF on the transcriptional activities of variants of the D4ex3 VNTR of the DRD4 genes of *P. troglodytes* and *H. sapiens* (Figure 5.7) showed that both TFs repressed the function of the *P. troglodytes* (PtD4ex3) VNTR construct. Conversely, only CTCF regulated the transcriptional activity supported by *H. sapiens* VNTR (HD4ex3) in cortical cultures; similar results were observed when the HD4ex3 were transfected into JAr cells (Figures 5.7 and 5.8).

I then tested if CTCF overexpression would have a regulatory effect on two common variants of the 5' promoter VNTR (D4L and D4S) found in the DRD4 gene of *H. sapiens* (Figure 5.9). The results show that the although under basal conditions the D4L and D4S promoter VNTRs did not support transcriptional activity, overexpression of CTCF induced significant repression of the pGL3p reporter gene expression. These results suggest that the CTCF protein interfering with the transcription initiation in the pGL3p, but only when the VNTR with a sequence that can potentially bind to CTCF has been introduce in the pGL3p vector. This may be caused by the recruiting effects CTCF has on the RNA polymerase II as proposed by Chernukhin (2007). The regulatory effects of these TF demonstrated that both 5' promoter and exon 3 VNTRs elements located in the DRD4 gene could be subject to regulation by CTCF in the CNS and perhaps suggest that the D4ex3 VNTR may act as a repressor of DRD4 gene expression.

The last section of this chapter analysed the regulatory effects of CTCF and YB-1 on two variants of a VNTR located in the 3' UTR of the SLC6A3 gene (DAT 9 and DAT 10) that also shared homology with the STin2 VNTR (Figure 1.12). In these experiments both CTCF and YB-1 proteins dramatically repressed the transcriptional activities of DAT 3' UTR 9 and DAT10 constructs (Figure 5.6b). The results

presented in this chapter corroborate that CTCF and YB-1 are versatile proteins, whose motifs are variable DNA consensus sequences and demonstrate that their regulation may affect the function of VNTRs in the SLC6A4, DRD4 and SLC6A3 genes in the CNS.

YB-1 and CTCF expression has been found to be responsive to stress signals in cellular environments with downstream effects on their regulation targets (Kohno et al., 2003; Ohlsson et al., 2001; Roberts et al., 2007). Thus, the activation/repression of the VNTRs function induced by modulation of YB-1 or CTCF ratio of absolute levels offers support to the hypothesis that the transcriptional activation of the SLC6A4, SLC6A3 and DRD4 genes in primates may be achieved through the stimulus-specific enhancer/repressor elements found in their VNTRs which are medating a reponse to stress (Bowen et al., 2005; Erblich et al., 2004).

In conclusion, the results presented in this chapter support the hypothesis that VNTRs could mediate a response to stress in the CNS, required for adaptability to novel challenges as presented when an organism adapts to new environmental conditions, which in turn may be mediated by TFs such as YB-1 and CTCF. However I postulate that in addition to these 2 factors, many others are likely to be involvd and currently this is an active area of research in the J Quinn group at the University of Liverpool.

Chapter 6 Non-coding evolutionary conserved regions in the DRD4 and SLC6A4 genes act as transcriptional regulators

6.1 Introduction

The identification of domains that have evolved to regulate neurotransmission of behaviour in the brain would contribute to a better understanding of the evolution of animal and particularly human behaviour and cognition. Multiple alignments of orthologous vertebrate genomes have permitted the discovery of non-coding evolutionary conserved regions (termed in this thesis as ECRs), typically embedded in large intergenic or intronic regions which can be as conserved as exons or as proximal 5' promoters. It has been proposed that these non-coding ECRs potentially exhibit *cis* regulatory activity, perhaps aided by chromatin modification processes (looping, tracking) which approximates these enhancers to their target promoter (Dean, 2006). The application of genome wide comparisons has allowed the discovery of ECRs in many structurally important genes (Mage et al., 1989; Prabhakar et al., 2006b; Shashikant et al., 2007), also in genes involved in neuronal development and behaviour (Davidson et al., 2006; Prabhakar et al., 2006b). Since modulation of neurotransmission during the evolution of *H. sapiens* have been correlated with the evolution of its cognitive abilities (Rockman et al., 2005), it is possible that changes in ECR function may have contributed to the evolution of the cognition of *H. sapiens.* In this chapter, I investigated the presence of ECRs in DRD4 and SLC6A4 genes and the evolution of an ECR in the DRD4 gene, and its ability to act as a *cis* regulator of transcription.

6.2 Aims

The initial aim of this chapter was to identify ECR regions with potential *cis* regulatory activity in the DRD4 and SLC6A4 genes loci. For this purpose, I analysed

the sequences of these two genes and their flanking found in modern humans and other mammalian and vertebrate species using the UCSC BLAT and ECR genome browsers.

The second aim of this chapter was to address *in vitro* the potential transcriptional activity of an ECR identified in the DRD4 gene (D4ECR1). As proof that these ECRs can affect gene expression, I cloned the ECR fragment into a reporter gene construct and measured its capacity to support reporter gene expression in dissociated cultures of neonate rat brain. Furthermore, using bioinformatics tools (Alibaba 2.1) I identified potential binding sites for TF in the D4ECR1 fragment and tested the regulatory effects of one of the identified TFs (Sp1) on the *cis* regulatory activity of the D4ECR1 *in vitro*. For this study I co-transfected the ECR construct with a human Sp1 expression vector *in vitro* into cortical cultures.

The third and final aim of this chapter was to identify changes in the sequence of D4ECR1 that has appeared during the evolution of primate lineages leading to *H. sapiens,* which may have contributed to the evolution of the human brain phenotype. For this purpose, I compared the sequences of the D4ECR1 of primates (*H. sapiens, P. troglodytes* and *M. mulatta*) and other mammals (*M. musculus, R. norvegicus* and *C. familiaris*) and reconstructed the evolution of this ECR in mammals using a phylogenetic approach, based on the TFBS found within their sequences.

6.3 Results

6.3.1 Identification of ECRs

To address whether ECRs were present in the DRD4 and SLC6A4 genes loci an *in silico* analysis of the loci and flanking regions was conducted using the ECR browser and the BLAT search of the UCSC browser. I set the ECR browser to detect sequences with a threshold of minimal length: 100 bp and minimal identity to human

orthologous sequences: 70%, which are the default parameters of the ECR browser. These parameters were used following Davidson (2006) which showed that these permit the identification of several functional ECRs in the PPTA gene loci. However, it is worth noting that regulatory domains vary greatly in size, thus the parameters used in this analysis are arbitrary. The validity of the identified ECRs was confirmed by analysing the same gene loci using the conservation tool of the UCSC BLAT search browser. This search produced an estimate of sequence conservation by comparing the human sequence to other mammal and vertebrate sequences (when available).

Analysis of the SLC6A4 gene showed that most of its intronic sequences exhibit great conservation (70% or more) of across mammals. Thus, to prioritise the identification of sequences with the highest degree of conservation, the ECR browser parameters were set to a more stringent level. Thus, the ECR browser identified sequences with a minimal of 85% conservation/100 bp length between distantly related mammals (*H. sapiens* [modern human]- *D. domestica* [marsupial]) or amongst at least three classes of mammals (e.g. *H. sapiens– C. familiaris-M. musculus*). This comparison revealed the presence of ECRs in the DRD4 and SLC6A4 genes (Figures 6.1 and 6.2).

6.3.1.1 ECRs in the DRD4 gene

Examination of the DRD4 gene (3.4 kb) and flanking regions (limited by the neighbouring genes at 2 kb upstream and 2.5 kb downstream) with the ECR browser identified three possible ECRs. Two ECRs were located in the first intron and the last one was present in the 3' flanking region. The first ECR (ECR1, 197 bp long, 93.5% conservation, Figure 6.1), located in the first intron was identified in *H. sapiens*, *M. mulatta* and *P. troglodytes.* This sequence is also present in the genomes of *R.*

norvegicus, M. musculus and *C. familiaris* (appendix 12.1) however, in the sequence conserved is shorter than in primates (71bp). As this element is shared by primates, rodents and carnivores, it is likely that it appeared in the DRD4 gene of mammals at least 80 mya, time calculated for the split of these mammalian taxa (Murphy et al., 2001). This is before the branches which originated these species (superorders Euarchontoglires: rodents, primates and insectivores and Laurasiatheria: e.g. carnivores, ungulates) separated (Figure 3.15 a and b).

The second ECR located in first intron (Figure 6.1; ECR2, 165 bp, 93.4% of conservation between *H. sapiens* and *M. mulatta,* appendix 12.1). This ECR was also identified in the genomes of *M. musculus* by the ECR browser. When verifying the authenticity of this ECR with the UCSC browser, it was observed that this area was not very conserved in rodents (45% conservation); and that this conservation increased in closer proximity to exon 2. Thus, it is possible that its reduced mutation rate in this region is caused by the increase purifying selection pressures which slow the exonic sequence from changing, and not because of its regulatory potential. Using the conservation tool of the UCSC BLAT search, the length of the conserved sequence reduced to 70 bp long, which was shorter than the set minimum of 100 bp (Davidson, 2006). The intronic sequences of genes from non-human species could be often incomplete, thus it is possible that these discrepancies appear for differences in the availability of intronic sequences between these genome browsers (Figure 6.1).

The ECR browser identified another potential ECR in the 3' region of *H. sapiens* and *C. familiaris* (ECR3, 70.5% conservation; 129 bp 0.8-1 kb from translation start site). Unexpectedly, this ECR was not identified in the DRD4 genes of species more closely related to humans as *M. musculus* or the *M. mulatta* (appendix 12.1). Examination of this region using the UCSC BLAT search browser

demonstrated that this sequence was not available for the rhesus macaque, and the degree of conservation found between *H. sapiens* and *M. musculus* sequences was relatively low (68%) in comparison to other ECRs idetified in this gene. It is possible that similar to the observed for the ECR 2, incomplete sequences of the 3' UTR region of these species could be responsible for the lack of conservation between closely related species such as *H. sapiens* and *M. mulatta*. It was therefore decided to concentrate on the first ECR identified in the DRD4 gene (D4ECR1) in future experiments.

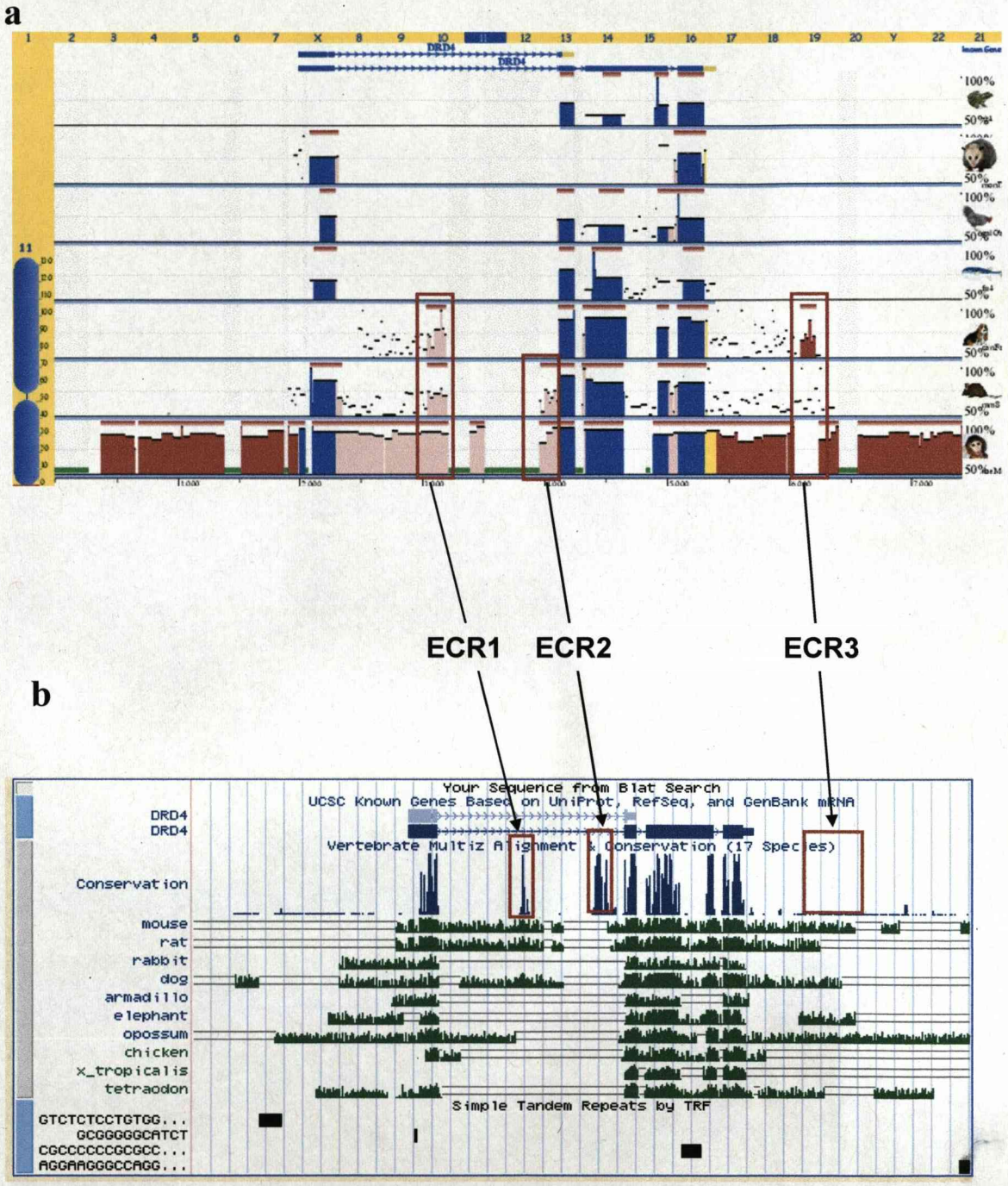


Figure 6.1 ECRs in the DRD4 gene of vertebrates. Two ECRs (ECR1 and ECR2) were identified in the first intron of the DRD4 gene of *H. sapiens*, *M. mulatta* and *M. musculus* when the ECR browser was set to identify conserved regions with a minimum of 70% conservation (a). ECR3 was not found by the UCSC BLAT broswer (b). The ECRs are marked by a red rectangle in (a) and (b).

6.3.1.2 ECRs in the SLC6A4 gene

I analysed the SLC6A4 gene (38kb approx.) and its flanking regions (limited by the neighbouring genes at ≈10 kb upstream and ≈4 kb downstream) of *H. sapiens* and other vertebrate species (Figure 6.2). The analysis of this locus using the ECR browser (minimal conservation 85%, 100 bp minimal length) demonstrated the presence of at least 7 peaks of conservation located in intronic regions and in the 5' and 3' region of the SLC6A4 gene of mammals (Figure 6.2). The analysis of the sequences using the conservation tool of the UCSC BLAT browser identified the presence of only 2 ECRs between *H. sapiens* and *Monodelphis domesticus* (modern humans). With the exception of the exonic regions, both genome browsers identified a sequence with the highest peak of conservation (conserved between *G. gallus* and *H. sapiens*) located in the 5' flanking region that corresponds to the 5' promoter of this gene (Figure 6.2).

The first ECR (ECR1) was a 115 bp sequence located in intron 12, conserved in 83.5% between *H. sapiens* and *M. domestica* (appendix 12.2). The second ECR (ECR2) was a sequence 254 bp long located in intron 1 (Figure 6.2). This ECR2 was 85.8 % conserved between *H. sapiens* and *C. familiaris* (appendix 12.2). Examination with UCSC BLAT browser confirmed that a similar (but shorter) sequence to ECR2 was also found in the serotonin transporter gene of *M. domestica* (appendix)

It is possible that differences in the availability of sequences in both genome browsers could be responsible for these inconsistencies. The degree of conservation presented by these two ECRs would indicate that they appeared in the SLC6A4 genes of mammals around 120-130 million years ago, before the ancestors which originated

marsupials and placental mammals (including humans and rodents) separated (Luo et al., 2003).

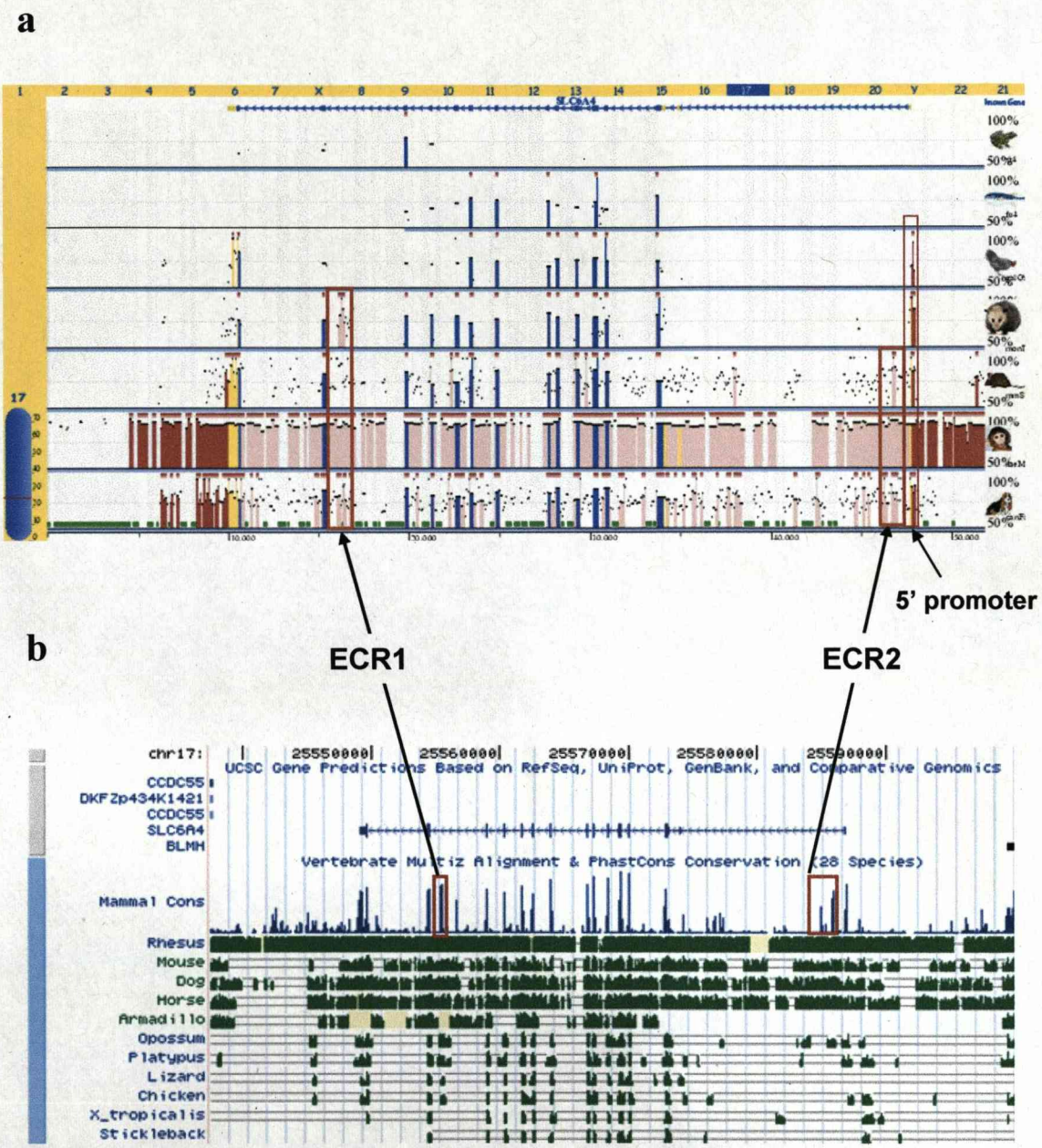


Figure 6.2 ECRs in the SLC6A4 gene of vertebrates. Two sequences were found conserved between the intron 1 (ECR1) and intron 12 (ECR2) of the SLC6A4 genes of *H. sapiens* (modern human) and *M. domestica* (opossum) by the ECR browser (a). The ECR browser was set to identify (in different vertebrate genomes) ECRs in with 85% homology with the *H. sapiens* sequence. The presence of these two ECRs was confirmed using the UCSC BLAT browser (b). In a, the 5' promoter of this gene was identified as a high peak of conservation across *G. gallus* (chicken) and *H. sapiens*. The ECRs are marked by a red rectangle in (a) and (b).

6.3.2 Functional analysis of the D4ECR1

6.3.2.1 The human D4ECR1 gene supports differential luciferase expression in cortical cultures of neonate frontal cortex

To investigate if the D4ECR1 identified in the DRD4 gene gene of mammals (Figure 6.1) had the ability to act as a modulator of gene expression in the developing mammalian CNS, I amplified and cloned the *H.sapiens* D4ECR1 into pGL3p and analysed its capacity to modulate reporter gene expression *in vitro*. The D4ECR1 construct was transfected into dissociated cultures of neonate rat frontal cortex, obtained as previously described (2.2.15.2) from Wistar rats (2 and 5 day old). Transfections were conducted using the protocol described in section 2.2.16.1.

The expression assays validated the predicted enhancer role of the human D4ECR1. This construct had the ability to act as an enhancer of pGL3p reporter gene expression in primary cultures obtained from 2 to 5 day old male Wistar rats (Figure 6.3). Its transcriptional activity varied from 6.4 fold (over the activity of pGL3p alone in 2 day old rats) to 3.15 fold increase (in 5 day old rats). The difference in fold increase supported in the two groups of rat pups was statistically significantly (Students' *T*-test, p=0.01, Table 6.1). These findings suggest that this ECR element can act as an enhancer of gene expression during neonatal development of the CNS, and perhaps contributes to the *cis* regulation of the DRD4 gene during development. Since the activity of the D4ECR1 construct was sensitive to the age of the cells, this assay further suggests the activity patterns of this element could be involved in the timing of the expression of the DRD4 gene, as suggested for ECRs in other genes (e.g.Prabhakar et al., 2006a). This observation deserves to be validated and extended by further *in vivo* work.

Figure 6.3

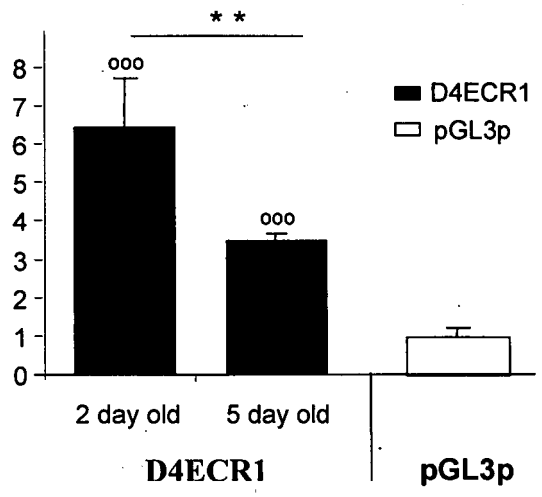


Table 6.1

	D4ECR1 activity		
animals age	2 days	5 days	pGL3p
average activity	6.403 (*) ooo	3.24 ooo	1.000
	±	±	±
SE	1.297	0.284	0.233

Figure 6.3 *H. sapiens* D4ECR1 exhibit transcriptional activity in dissociated cultures of neonate rat cortex. The D4ECR1 construct (1μg) was transfected into dissociated cultures of frontal cortex obtained from 2 and 5 days old male wistar rats under basal conditions using ExGen 500. Student's *T*-test showed that the D4ECR1 construct induced significant reporter gene expression different from pGL3p alone, (where ooo indicate $p \leq 0.005$).

Table 6.1 The transcriptional activity of the D4ECR1 in cultures derived from 2 and 5 day old rats is significantly different. Statistical analysis showed that the difference in function of the D4ECR1 construct found in the 2 and 5 day old cultures were be significant (where * indicate $p \leq 0.05$; values obtained in three independent experiments per triplicate (n=9).

6.3.2.2 *In silico* analysis of the D4ECR1

In order to investigate which TFs were potentially interacting with the D4ECR1 found in mammals (Figure 6.1), the ECR sequences were subject to an *in silico* analysis, using the publicly available Alibaba 2.1 program. The programs detected known consensus binding sequences for TF based on the TRANSFAC database (4.0).Alibaba 2.1 detected binding sites for TF (TFBS) using the parameters: described previously (section 2.2.20.1).

The TFBS identified by the program Alibaba 2.1 are shown in Table 6.2. Briefly, the Alibaba 2.1 program identified binding sites for 19 different types of TFs in the ECR1 of *H. sapiens*, 22 sites in *M. mulatta* and *P. troglodytes*; 19 and 16 sites in the *M. musculus* and *R. norvegicus* and 20 sites in the sequence of *C. familiaris*. Binding sites conserved across all ECR1 sequences included: Sp1, c-jun, CREB, CPE-bind, ATF, CRE-BP1 and C/EBPalp. The binding site for Sp1 was the most common and it was found in every ECR sequence examined. Interestingly, the ECR of *H. sapiens* had the lowest number of different types of TFBS amongst the primate ECR sequences analysed (Table 6.2).

Sp1 has been implicated in the regulation of dopamine systems in the rat brain (Zhou et al., 2004) and has been reported to be expressed in the frontal cortex of neonates of rats (Hood et al., 2000). Moreover, Sp1 has been suggested to modulate in the regulation of the human DRD4 gene via other *cis* regulators present elsewhere in this gene (Ronai et al., 2004). The presence of binding sites for Sp1 suggested that this TF could be a good candidate for regulating the expression of the DRD4 gene via binding of the D4ECR1 enhancer. Therefore, in the following section I investigated the regulatory effects of the overexpression of Sp1 on this D4ECR1 of *H. sapiens* in

primary cultures of neonate rat cortex which express the DRD4 gene (appendix 2 and Araki et al., 2007).

	H. sapiens	*P. troglodytes*	*M. mulatta*	*M. musculus*	*R. norvegicus*	*C. familiaris*
GR	2	2	0	3	2	0
SP1	5	6	6	3	5	11
ETF	1	0	1	0	0	2
USF	1	1	1	0	0	0
REV-ErbA	2	1	0	0	2	0
COUP	1	1	1	1	0	0
ER	1	1	1	1	0	2
c-Jun	1	1	2	1	2	1
CREB	1	1	1	1	1	1
CPE-bind	1	1	1	2	1	1
ATF	1	1	1	1	1	1
E1A-12S	1	1	1	0	0	1
CRE-BP1	1	1	1	1	1	1
C/EBPalp	1	1	1	1	1	1
ADR-1	1	1	1	0	0	0
AP2 alp	1	0	0	1	1	1
Oct-1	1	1	1	2	2	0
myogenin	1	1	0	0	0	0
AP1	2	2	1	1	1	0
T3Ralp	0	1	1	0	0	0
RAR-alp	0	1	1	0	0	1
RXRalp	0	1	1	2	1	1
GCN4	0	1	1	1	1	0
Pu-1	0	1	1	0	0	0
Jun-D	0	0	1	0	0	0
repressor	0	0	1	0	0	0
Ftz	0	0	0	1	0	0
PR	0	0	0	2	0	0
T3R	0	0	0	2	0	0
Ttx	0	0	0	1	0	1
RXR beta	0	0	0	0	1	0
c-fos	0	0	0	0	1	0
Egr-1	0	0	0	0	0	2
Adf 1	0	0	0	0	0	1
NF	0	0	0	0	0	1
NF-muE1	0	0	0	0	0	1
Tra1	0	0	0	0	0	1
T3R beta	0	0	0	0	0	1

Table 6.2 Comparison of putative TFs binding to the D4ECR1 sequences identified by Alibaba 2.1. In the D4ECR1 sequences the Sp1 sites were most abundat. The number of sites for Sp1 are highlighted in red font in the table.

6.3.2.3 Sp1 regulates the D4 ECR1 enhancer activity in primary cultures of neonate rat frontal cortex.

To investigate if Sp1 modulates the D4ECR1 ability to support reporter gene expression as inferred from *in silico* analysis, (Table 6.2), I tested the effect of the over-expression of Sp1 on the enhancer activity of the D4ECR1. I co-transfected the D4ECR1 construct (1μg per well) with two concentrations (0.5 and 1μg per μg of Sp1 construct) of an expression vector carrying the full length cDNA of human Sp1 into primary cultures of neonate rat frontal cortex obtained from 2 days old rats as previously described in section 2.2.15.2. The unmodified luciferase vector pGL3p was also co-transfected with the two concentrations of Sp1 (as described in section 2.2.17) to control for possible effects on the backbone of the luciferase plasmid.

This co-transfection experiment demonstrated modulation of the D4ECR1 enhancer acting by Sp1 (Figure 6.4). The overexpression of Sp1 (both concentrations) had a repressing effect on the transcriptional activity of the D4ECR1 construct (Table 6.3). Sp1 also repressed the activity of pGL3p although this was found to be not significant (Table 6.3). Thus, I normalised the basal transcriptional activity of the D4ECR1 and pGL3p to 100% and expressed the values of the co-transfection with Sp1 as percentages of repression. These comparisons showed that Sp1 induced repression of the basal transcriptional D4ECR1 construct (83.5% and 77.5% repression with 0.5 and 1 μg of Sp1) and this effect was statistically significant (Table 6.3). These results suggest that the D4 ECR1 could be regulated by Sp1 *in vivo* in the CNS.

Figure 6.4

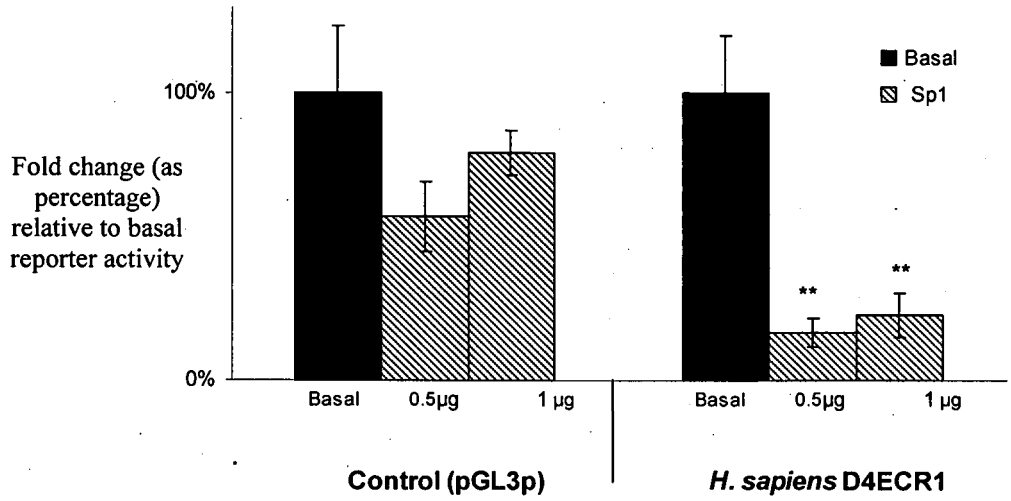


Table 6.3

	pGL3p			D4ECR		
	Basal	**0.5µg Sp1**	**1µg Sp1**	**Basal**	**0.5µg Sp1**	**1µg Sp1**
Average	1.001	0.570	0.791	1.000	0.165 **	0.225**
	±	±	±	±	±	±
SE	0.233	0.123	0.078	0.203	0.049	0.077

Figure 6.4 Regulatory effects of Sp1 on the activity of the D4ECR1 in disassociated cultures of rat frontal cortex. Two concentrations of Sp1 (0.5 and 1µg) were co-transfected with 1 µg of the D4ECR1 construct (activity of the D4ECR1 construct co-transfected with Sp1 is represented by white bars with oblique black lines).

Table 6.3. Sp1 significantly downregulated the reporter gene expression supported by D4ECR1 but not of pGL3p. Statistical analysis shows that 0.5 and 1 µg of Sp1 singnificantly repressed the transcriptional activity of D4ECR1 (Student's T-test, where ** indicates p=0.01; *n*=9)

6.3.2.4 Comparison of reporter gene expression supported by the D4 ECR 1 in different regions of the rat neonate brain

The option to generate transgenic animals to address the D4ECR1 was not available to me at the time of this work. To address tissue-specific expression of this D4ECR construct in CNS, as demonstrated for other ECR enhancers (Davidson et al., 2006; Pennachhio et al., 2006), I transfected the D4ECR1 construct into primary cultures of frontal, temporal cortex and midbrain areas which have been reported to differ in endogenous DRD4 expression levels (Gong et al., 2003). Primary cultures of these three areas were obtained as described previously (section 2.2.15.2), from 5 day old male Wistar rats.

These experiments did not demonstrate any difference in the tissue specificity of the activity of the D4ECR1 enhancer in the neonate rat brain (Figure 6.5). Although the D4ECR1 construct was able to support reporter gene expression in all cultures (supporting 3.24, 3.18 and 3.42 fold increase in the frontal, temporal cortex and midbrain respectively) the levels of activity were not statistically different (Students' *T*-test, frontal cortex vs. midbrain p= 0.34; frontal vs. temporal cortex p=0.44 and midbrain vs. temporal cortex p=0.27). The tissue specificity of this regulatory element in the brain cannot be discarded until a transgenic model is prepared.

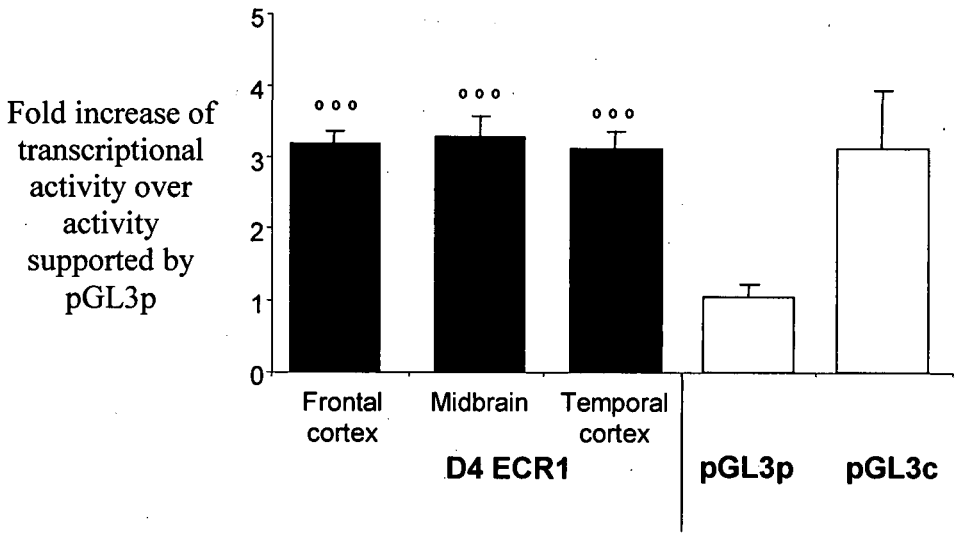


Figure 6.5 Transfection of the *H. sapiens* D4ECR1 into disassociated cultures of rat frontal cortex, temporal cortex and midbrain. The D4ECR1 construct supported significant levels of reporter gene expression in the three types of tissue, but the levels supported were not significantly different (Student's *T*-test p=0.34 [Frontal cortex vs. Midbrain]; p=0.44 [Frontal cortex vs. Temporal cortex] and p=0.27 [Midbrain vs. Temporal cortex]). The levels of reporter gene expression supported by the D4ECR1 construct were equal to the positive control (pGL3c). The transfections were conducted in cultures derived from 5 day old neonates, *n*=9).

6.3.3 Evolution of the D4ECR1

6.3.3.1 Alignment of the D4ECR1 sequence of primates

In this section, I investigated which changes had occurred in the D4ECR1 sequences in the primate lineages leading to modern humans. For this purpose, I aligned sequences of the D4ECR1 of 3 primates: *H. sapiens*, *P. troglodytes* and *M. mulatta* using the clustalW tool of the EBI server (http://www.ebi.ac.uk/Tools/clustalw/).

The alignment produced is shown in Figure 6.6. The tool aligned the sequences of primates searching for the best fit. This analysis also demonstrated the presence of 8 sites at which the primate sequences differed. In brief, these changes were caused by 7 substitutions and 1 indel (Figure 6.6). The sequence of the *H. sapiens* exhibit 3 unique mutational events: 2 substitutions and the lack of 1 base or indel (in red font Figure 6.6), the sequence of *P. troglodytes* presented 2 species-specific substitutions (in blue font, Figure 6.6) and 3 substitutions were exclusive to *P. troglodytes* and *H. sapiens* (in green font Figure 6.6). The sequence differences identified in the *H. sapiens* and *P. troglodytes* D4ECR1 potentially create differences in the type of TFBS found in their D4ECR1 (Table 6.2). For example, the change of C (seen in the D4ECR1 sequences of *P. troglodytes* and *M. mulatta*) for a G in the sequence of *H. sapiens* transformed a putative binding site for T3R/RXR/RAR- α into a Sp1 and REV-ErbA site (highlighted by a dashed-line rectangle in Figure 6.6). Similarly, the change of a T for a C in the ECR sequence of *H. sapiens* created a TFBS for AP2-α (Figure 6.6). The loss of a retinoic acid binding site (e.g. RAR or RXR) in the modern human D4ECR1 has potential relevance in the regulation of the DRD4 gene, as the retinoic acid is a known regulator of dopaminergic genes during development (Krezel et al., 1998). However, it is noteworthy that this D4ECR1

variation present 2 SNPs which affect the putative TFBS found in the VNTR. Indeed, one SNP (C/T) affect the AP2α site unique to the *H. sapiens* D4ECR1 (indicated by an arrowhead in Figure 6.6) and a second SNP (C/T) could transform a site for REVErbA (indicated by an arrowhead in Figure 6.6) into a site for RARα, RXRα and T3Rα. Therefore, the potential differences these SNPs could create may affect the D4ECR1 regulatory properties in modern human populations. Thus, more individuals of each species analysed in this study should be studied to confirm whether there is a inter-specific variation in the D4ECR1 sequences.

The D4ECR1 of *H. sapiens* fragment was demonstrated to support robust reporter gene expression in cultures derived from neonatal CNS tissue (Figure 6.3); therefore, it is possible that the sequence differences identified, which have emerged during the evolution of *P. troglodytes* and *H. sapiens*, potentially correlate with differential regulation of the DRD4 gene *in vivo* between these two species. This observation deserves further investigation *in vivo* in future experiments.

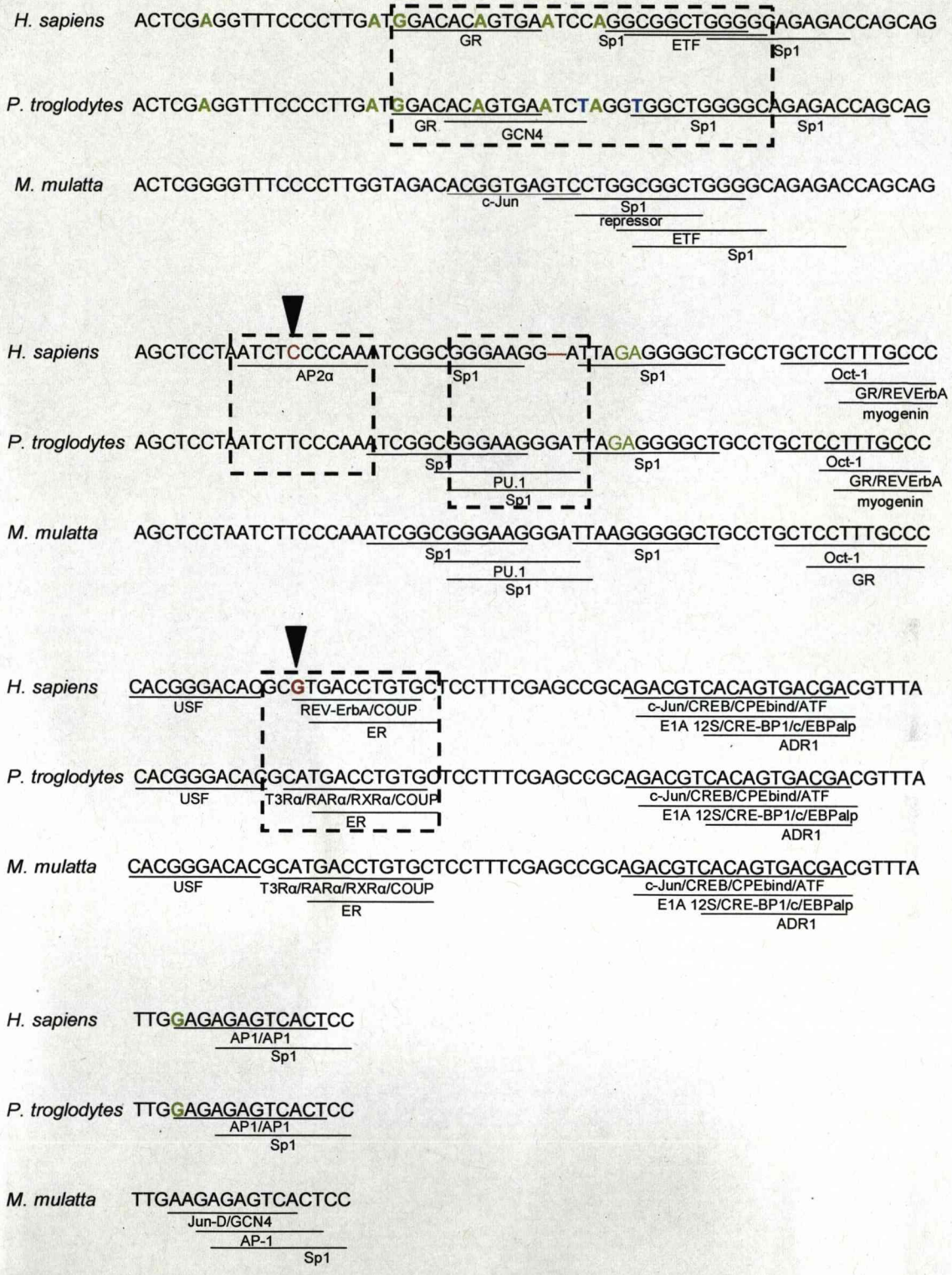


Figure 6.6 Alignment of the D4 ECR1 of 3 primate species. This alignment shows D4ECR1 sequences (239 bp) found in the first intron of the DRD4 gene of *P. troglodytes* and *H. sapiens* and *M. mulatta*. In the sequences nucleotides in blue font are changes exclusive to the *P. troglodytes* sequence, in red font are changes exclusive to *H. sapiens* and in green font are changes exclusive to the ECRs of *H. sapiens* and *P. troglodytes*. The potential TFBS identified by Alibaba 2.1 were aligned to the sequences. Differences between the TFBS in *H. sapiens* and *P. troglodytes* ECRs are marked by dash-line rectangles.

6.3.3.2 Evolution of the D4ECR1 of mammals based on TFBS

In this section, I analysed how the D4ECR1 sequences had changed during evolution of mammals. For this study, I conducted a phylogenetic analysis based on the TFBSs identified within the sequences (using Alibaba 2.1) of the D4ECR1 sequences of these mammalian species (Table 6.2). The matrix is shown in appendix 13 and the cladogram produced is shown in Figure 6.7.

In the cladogram, the *C. familiaris* sequences are located at the base of the tree, followed by a branch that bifurcates into two: a rodent branch (formed by the sequences of *R. norvegicus* and *M. musculus*) and a primate branch (formed by the sequences of *H. sapiens*, *P. troglodytes* and *M. mulatta*). It is noteworthy that even though the sequences of rodents and primates share similarities, which cluster them in their respective branches, each D4ECR1 sequence is distinct, as reflected by the length of the branches (BL). The BL of the rodent D4ECR1 is similar to the observed in the primate branch. This would suggest that the D4ECR1 in rodents is evolving at a similar rate as the ECR of primates. Interestingly, the cladogram also shows that *M. musculus* has accumulated more changes than *R. norvegicus* (BL =5 and 4.5 = respectively). In the primate branch, the *M. mulatta* sequence occupies the basal branch to *P. troglodytes* and *H. sapiens*. It is important to notice that in this latter branch the BL of *P. troglodytes* is equal to 0, while this value is equal to 2 in *H. sapiens*. This indicates that most of the evolutionary changes occurred between these species have occurred in the *H. sapiens* D4ECR1.

In conclusion, this cladogram shows that the sequence of the D4ECR1 of the DRD4 gene has evolved in a similar fashion as sequences undergoing purifying selection (e.g. exonic sequences, which may reflect the great conservation exhibited by the D4ECR1). Nevertheless, the cladogram also demonstrates that these sequences

could rapidly accumulate variation as seen in the rodent D4ECR1 branches and in *H. sapiens* branch, which may indicate specific evolutionary forces, which permit the rapid turnover of the binding sites of these ECRs in these species, perhaps as a response to their specific environmental needs.

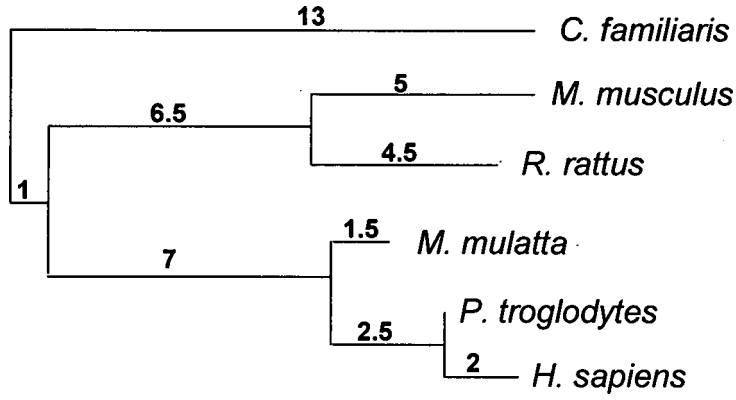


Figure 6.7 Cladogram of the D4ECR1 based on TFBSs. The TFBS found in the ECR sequences have varied more between the two rodent species (BL=9.5) than amongst the 3 primate species (BL=6). In the clade formed by the sequences of *P. troglodytes* and *H. sapiens*, the *H. sapiens* D4ECR1 has accumulated more changes affecting TFBS than *P. troglodytes* ECR sequence (as indicated by their BL, being 2 for the *H. sapiens* branch and 0 for the *P. troglodytes* branch).

6.4 Brief summary and discussion

This chapter aimed to identify and analyse the potential role of non-coding evolutionary conserved sequences as potential *cis* regulators of the DRD4 and SLC6A4 genes. The analysis demonstrated that there are several conserved domains in non-coding regions of these two genes (Figure 6.1 and 6.2). In some cases these non-coding ECRs have been conserved during the evolution of all mammals (e.g. ECR1 in the SLC6A4 gene), also there are ECRs which have only appeared more recently, during the evolution of placental mammals (e.g. ECR1 of the DRD4 gene). This suggests that the variation could modulate the *cis* regulation of the expression profiles of these two genes in mammals.

In section 6.3.2.1 I focused on analysing *in vitro* the potential regulatory role of one of the D4ECR1 identified in the DRD4 gene in section 6.3.1.1. The first analysis showed that in dissociated cultures of neonate rat cortex, the D4ECR1 supported high levels of reporter gene expression. Interestingly its transcriptional levels varied depending on the age of the animals used for the experiments. These results would be consistent with a proposed *cis* regulatory role of this ECR in the developing CNS, and furthermore, suggest temporal regulation of the transcriptional activities of this ECR in the CNS. This latter hypothesis needs further study *in vivo* perhaps using transgenic models, which have been successfully used to demonstrate this property of other *cis* regulatory ECR domains (Mackenzie and Quinn 2004, Davidson et al., 2006, Prabhakar et al., 2006a).

In section 6.3.2.2 I focused in the identification of potential TFs, which might mediate the regulatory activity of this D4ECR1 observed in cortical cultures. This was achieved by the use Alibaba 2.1. This program revealed the presence of several TFBS in the D4ECR1 that could potentially modulate its activity. The most common type of

binding site found in the ECRs of the mammals analysed was for the TF Sp1. Indeed, the co-transfection experiments described in section 6.3.2.3 validated the regulation of Sp1 on the transcription activity of the D4ECR1 of *H. sapiens* in the CNS *in vitro*. This results offers support to the hypothesis that Sp1 mediate the expression of the *H. sapiens* DRD4 gene proposed by *in silico* and *in vivo* studies. For example, *in silico* studies have demonstrated the presence of Sp1 sites proximal to and on the regulatory region of this gene (Seaman et al., 1999; Szantai et al., 2005; D'Souza et al., 2004 and Kamakura et al., 1997). Furthermore, the interactions of Sp1 with the two variants of the promoter VNTR have been demonstrated by capillary electrophoretic mobility shift assay (CEMSA) which is a modified EMSA assay for small samples (Ronai et al., 2004). Since binding sites for Sp1 were found in each species D4ECR1 analysed, the results suggest that the function of this conserved region could be mediated by Sp1 *in vivo*. The results also offer support to the use of bioinformatics tools for the identification of potential interactions between regulatory domains and TF, however, these need to be verified *in vitro* and whenever possible *in vivo*.

The final section of this chapter (section 6.3.3) investigated the following question: If this D4ECR1 can regulate the expression pattern of the DRD4 gene, which is linked to behaviour and cognitive abilities; are there any sequence changes that have occurred in the history of *H. sapiens* that may be correlated to the evolution of the DRD4 expression pattern? The results presented in section 6.3.3.1 suggested that there are indeed mutations exclusive to the *H. sapiens* D4ECR1 that have accumulated during its evolution after it split from the apes. In comparison to other primates D4ECR1, such changes potentially create and eliminate TFBS for several TFs, which have been proposed to play an important role in the regulation of dopaminergic neurogenesis during mammalian embryonic development (Krezel et al.,

1998). However, the functional significance of these changes remains to be demonstrated in gene expression studies of the *P. troglodytes* and other non-human primates D4ECR1 in the future. Nevertheless, these preliminary results suggest that changes in the D4ECR1 may have contributed to the evolution of the expression profile of the DRD4 gene during evolution, which may have contributed to the overall evolution of modern human cognition. Additionally, SNPs present in the *H. sapiens* D4ECR1 could affect the expression of the DRD4 gene in the CNS with consequences in behaviour and disease.

Chapter 7 General Discussion

In the recent years, it has become evident that changes in the *cis* regulation of gene expression play a key role in the phenotypic adaptability and evolution of behaviour at all organismal levels. Accordingly, recent studies comparing the patterns of gene expression of hominid brains suggest that the change of *cis* regulatory domains may have also participated in the behavioural and cognitive evolution of *H. sapiens* (Heissig et al., 2005; Prabhakar et al., 2006a). Consistent with this, I demonstrated in chapter 3 that changes in the *cis* regulatory domains located in DRD4 and SLC6A4 genes have occurred during the evolution of hominids and that these may alter gene expression that mediates changes in behaviour and cognition.

The findings in chapter 3 are consistent with the hypothesis that the sequence and evolutionary pathways of the variable number tandem repeats (VNTRs) found in the DRD4 and SLC6A4 genes that act as *cis* regulators of gene expression are distinct between *H. sapiens* and its closest living relatives, the *P. troglodytes* and *Gorilla* sp. For example, the analysis of the sequences of the exon 3 VNTR of the DRD4 (D4ex3) gene of hominids showed that the VNTRs of *P. troglodytes* and *Gorilla* sp. are markedly different from the VNTR of *H. sapiens* (Figure 3.21b). In the SLC6A4 gene, the STin2 VNTRs of *P. troglodytes* and *Gorilla* sp. follow a similar trend as that seen in the D4ex3 VNTR (Figure 3.10a), which is one of divergence between the TFBS found in modern human and great apes VNTRs. The reconstruction of their evolution based on transcription factor binding sites (TFBS) found in the VNTRs suggest that these VNTRs have evolved similarly by mechanisms of parallel evolution, and furthermore, that their evolutionary trajectories reflect functional plasticity. Importantly, mere random recombination processes cannot be the only mechanism to explain these data, as the D4ex3 VNTR is exonic whereas the STin2

VNTR is intronic and consequently the evolutionary forces applied on these regions would be typically distinct. Exons have been found to be greatly conserved across vertebrates and even invertebrates, while introns usually experience higher evolutionary rates.

The sequences of the 5' promoter VNTR found in the SLC6A4 gene of hominids are greatly conserved, and their evolution reflects less functional plasticity than that observed in the STin2 VNTR (Figure 3.13a vs. 3.10a). This difference indicates that the STin2 and 5'promoter VNTRs of the SLC6A4 gene are evolving at different rates, and this is perhaps caused by the proximity of the latter to the core promoter of the SLC6A4 gene. Due to the general conservation in promoter regions, this area may be under great constraint to remain constant.

Interestingly, the analysis of the sequences of the STin2 and D4ex3 VNTRs variants also show that the sequences of *H. sapiens* and *P. pygmaeus* share common traits (TFBS). Such similarity may not necessarily arise from similar behaviour as in the previous case, given the great time (10-14 million years) which has passed since *P. pygmaeus* and *H. sapiens* last shared a common ancestor. Nevertheless, recent studies on the behaviour of hominids suggest that some behavioural traits exhibited by early humans are shared with *P. pygmaeus* (Thorpe et al., 2007) in which case, it might suggest a case of convergent evolution to explain the similarities amongst their VNTRs.

The differences found in the sequences of the SLC6A4 (STin2 and promoter) and DRD4 (D4ex3) VNTRs between modern humans and great apes can affect the putative TFBS found in these domains. Consequently, this could correlate to distinction between the transcriptional profiles of the SLC6A4 and DRD4 VNTRs of *H. sapiens* and the other hominids (Raghanti et al., 2007). The *in vitro* analyses

conducted in chapter 4 are in agreement with this hypothesis. The STin2 VNTRs of *P. troglodytes* and *Gorilla* sp. exhibit lower transcriptional activity (Figure 4.4) than the *H. sapiens* STin2.9 and STin2.12. The average transcriptional activity of the third VNTR variant of *H. sapiens*, STin2.10, was higher than the VNTRs of *P. troglodytes* and *Gorilla* sp., but these differences were not statically significant. These results suggest a trend of variation amongst *H. sapiens* and *P. troglodytes* and *Gorilla* sp. STin2 VNTRs. It is possible that the variability inherent to primary neuronal cultures used in these experiments is in part responsible for this effect. In the case of the SLC6A4 promoter VNTR, the transcriptional activities of the variants were similar amongst hominids, with the exception of the *P. troglodytes* VNTR reporter gene construct (Figure 4.8b). This finding demonstrates that in spite of the greater similarity demonstrated by the promoter VNTR of hominids, even minimal variation in *cis* regulatory sequences could invoke a major change in their regulatory activities. Furthermore, since the *H. sapiens* STin2 and promoter VNTRs are likely to have synergistic effects on SLC6A4 gene expression, thus it is possible that the lack of activity of the *P. troglodytes* promoter VNTR could also affect the overall regulation of gene expression mediated by these VNTRs. This synergistic activity could result in differential regulation of the SLC6A4 gene expression between *H. sapiens* and its closest relative, the *P. troglodytes*; therefore, the additive or synergistic effects of all reported VNTR variants found in the SLC6A4 locus of *P. troglodytes* must be explored in future experiments.

The functional analyses comparing the most common D4ex3 VNTR variant found in *H. sapiens* and one common VNTR variant found in *P. troglodytes* demonstrated that sequence differences correlated with transcriptional differences *in vitro* (Figure 4.9b). Indeed, the activity of the *H. sapiens* D4ex3 VNTR construct was

active in three tissue culture models (JAr, SN4741 cells and primary cultures of neonate rat cortex) and exhibited tissue specificity (acting as a repressor in cortical cultures and as an enhancer in SN4741 and JAr cell cultures). Conversely, the *P. troglodytes* VNTR construct was not active. This preliminary study indicates distinct regulatory properties of these two variants of the D4ex3 VNTRs of *H. sapiens* and *P. troglodytes*. However, to assert whether the diverse variants of the D4ex3 VNTR could induce differential expression of the DRD4 gene between these species, more D4ex3 VNTR variants and other regulatory domains found in this gene locus should be studied *in vitro*.

The activity of STin2 VNTRs of modern humans is regulated *in vivo* and *in vitro* by YB-1 via Y boxes found in the VNTR sequences (Roberts et al., 2007 Klenova et al., 2004). The protein CTCF, a binding partner of YB-1, can also regulate YB-1 interaction with the STin2 VNTRs and directly through CTCF binding sites. As detailed in chapter 3, the sequences of the STin2 VNTRs of all hominids studied comprise Y-boxes and CTCF sites, therefore suggesting these VNTRs can also be modulated by YB-1 and/or CTCF. The findings of chapter 5 confirm this hypothesis, since overexpression of CTCF and YB-1 represses the activities of all primate STin2 VNTRs constructs tested in neuronal cultures (Figures 5.2 and 5.3). The number and type of Y boxes within the *H. sapiens* STin2 VNTRs correlated to differential binding of YB-1 for the 3 VNTR variants. A correlation between the number of Y-boxes and differential regulation by YB-1 has been reported for regulatory domains found in other genes e.g. collagen type I gene *COL1A1*, (Norman et al., 2001). Therefore, the variability of the number and type of Y boxes and the distance between these boxes found in the STin2 VNTR sequence of each primate species analysed suggest differential regulation by YB-1 and CTCF. However, the co-transfection experiments

conducted in chapter 5 showed that all STin2 VNTRs were equally down regulated by overexpression of YB-1 and CTCF, regardless of their sequence variation. It is possible that the luciferase assays are not sensitive enough to identify subtle differential regulation of YB-1 and CTCF on the different VNTRs. Therefore, potential differences in the *cis* regulatory activities of these *cis* acting domains should be analysed in transgenic animal models, which are important tools to identify spatial differences in their activities (Mackenzie and Quinn 1999, 2004; Prabhakar et al., 2006a)

The similarity in the sequence of the VNTRs of the DAT (3' UTR), SLC6A4 (promoter) and DRD4 genes (exon 3) to the STin2 VNTR, suggested that all these domains could be regulated in part by similar TFs. The co-transfections experiments detailed in chapter 5 (see Figures 5.6b, 5.7 and 5.8) support this hypothesis. In cortical cultures, overexpression of YB-1 and CTCF down regulated the basal transcriptional activities of all human and non-human primate STin2 and promoter VNTR constructs. A similar effect was observed for the DAT 3' UTR VNTRs of modern humans tested (DAT9 and DAT 10, Figure 5.6b). In the case of the D4ex3 VNTR of the DRD4 gene, the *H. sapiens* construct (HD4ex3) was down regulated by CTCF, however, YB-1 had no effect on its basal transcriptional activity. Meanwhile, the *P. troglodytes* VNTR construct (PtD4ex3) was down-regulated by both YB-1 and CTCF (Figure 5.7). This repression is not due to an effect on backbone (pGL3p, appendix 8) as the same constructs are used for other experiments carried out by our group where YB-1/CTCF activate these VNTR sequences in a similar manner (Klenova et al.,2004; Roberts et al., 2007).

The last section of this work focused on the study of non-coding evolutionary conserved regions (ECRs) in the DRD4 and SLC6A4 genes. Changes in the

regulation of the DRD4 or SLC6A4 gene expression that occur during early (prenatal or neonatal) development has great consequences in cognitive and behavioural phenotype in later life, as documented in humans and rats (Herlenius and Lagercrantz, 2001; Mabandla et al., 2007). It is likely that such changes in gene expression could be mediated by regulatory domain active during development; however, only few attempts to identify such elements in neurotransmitter genes have been conducted (Davidson et al., 2006; Pennacchio et al., 2006; Prabhakar et al., 2006b). The results of chapter 6 suggest the presence of a class of domains, which may act as transcriptional regulators during early CNS development in the SLC6A4 and DRD4 genes (Figure 6.1 and 6.2) of some placental (modern) mammals. These ECRs harbour abundant putative binding sites for a diversity of TFs, some of which have been linked to the regulation of embryonic development of the CNS (Table 6.2). Preliminary analysis of the *in vitro* transcriptional activity of the D4ECR1 showed that its activity correlates with the developmental stage of the animal used for the production of the cultures; being significantly higher in cultures obtained from younger than older animals (Figure 6.3). Although this data is preliminary, it has potential implications in the *cis* regulation of neurotransmission and therefore deserves future studies. It is possible that the transcriptional activity of this regulatory domain is tissue specific in the CNS, however, this remain to be tested in future experiments using a transgenic animal model.

Finally, the link of the DRD4 gene with cognitive function in humans suggests that there may have been a change in the patterns of expression of this gene during the evolution of *H. sapiens* from the expression pattern found in the ancestral ape. Comparisons of the sequences of the D4ECR1 of 3 primates (*M. mulatta, P. troglodytes* and *H. sapiens*) conducted in chapter 6 (Figure 6.6) offer support to this

hypothesis. This shows that 5 mutations/deletions distinguish the *H. sapiens* and *P. troglodytes* sequences analysed, with 3 of these mutational events having occurred exclusively in the D4ECR1 of *H. sapiens*. Such mutations can affect the putative binding of TFs to the ECR, notably eliminating sites for binding of retinoic acid, a known modulator of DRD4 expression in the brain (Krezel et al., 1998).

In conclusion, I propose that the findings of this study is that similar mechanisms modulating VNTRs and ECRs to mediate in part the divergence of the patterns of expression of the DRD4 or SLC6A4 genes between *H. sapiens* and its closest living relatives *P. troglodytes* and *Gorilla* sp. Nevertheless, the present study has demonstrated that, some transcriptional regulatory domains proposed to contribute to the diversification of the expression profile of modern human's brain indeed differ from those found in the African apes. This would suggest that the gene expression profile of the *H. sapiens* brain has changed to adapt to new environmental conditions from the ancestral ape stock. Furthermore, the variation observed in the regulatory domains of the DRD4 and SLC6A4 genes, together with many others (e.g. Table 4.8) could contribute to explain the evolution of the outstanding cognitive and behavioural phenotype of modern humans.

Bibliography

Aguirre AJ, Apiquián R, Fresán A, and Cruz-Fuentes C (2007) Association analysis of exon III and exon I polymorphisms of the dopamine D4 receptor locus in Mexican psychotic patients. Psychiatry Research 153: 209-215.

Allman J, Hakeem A, and Watson K (2002) Two phylogenetic specializations in the human brain. Neuroscientist 8:335-346.

Allman JM, Hakeem A, Erwin JM, Nimchinsky E, and Hof P (2001) The anterior cingulate cortex. The evolution of an interface between emotion and cognition. Ann N Y Acad Sci 935:107-117.

Alvarez C, Vitalis T, Fon EA, Hanoun N, Hamon M, Seif I, Edwards R, Gaspara P, and Cases O (2002) Effects of genetic depletion of monoamines on somatosensory cortical development. Neuroscience 115:753-764.

Andres AM, Soldevila M, Navarro A, Kidd KK, Oliva B, and Bertranpetit J (2004) Positive selection in MAOA gene is human exclusive: determination of the putative amino acid change selected in the human lineage. Hum Genet 115:377-386.

Araki KY, Sims JR, and Bhide PG (2007) Dopamine receptor mRNA and protein expression in the mouse corpus striatum and cerebral cortex during pre- and postnatal development. Brain Res 1156:31-45.

Arbelle S, Benjamin J, Golin M, Kremer I, Belmaker RH, and Ebstein RP (2003) Relation of shyness in grade school children to the genotype for the long form of the serotonin transporter promoter region polymorphism. Am J Psychiatry 160:671-676.

Asghari V, Sanyal S, Buchwaldt S, Paterson A, Jovanovic V, and Van Tol HH (1995) Modulation of intracellular cyclic AMP levels by different human dopamine D4 receptor variants. J Neurochem 65:1157-1165.

Bachir LK, Garrel G, Lozach A, Laverriere JN, and Counis R (2003) The rat pituitary promoter of the neuronal nitric oxide synthase gene contains an Sp1-, LIM homeodomain-dependent enhancer and a distinct bipartite gonadotropin-releasing hormone-responsive region. Endocrinology 144:3995-4007.

Bachner-Melman R, Dina C, Zohar AH, Constantini N, Lerer E, Hoch S, Sella S, Nemanov L, Gritsenko I, Lichtenberg P, Granot R, and Ebstein RP (2005) AVPR1a and SLC6A4 gene polymorphisms are associated with creative dance performance. PLoS Genet 1:e42.

Bailey JN, Breidenthal SE, Jorgensen MJ, McCracken JT, and Fairbanks LA (2007) The association of DRD4 and novelty seeking is found in a nonhuman primate model. Psychiatr Genet 17:23-27.

Bakewell MA, Peng S, and Zhang J (2007) More genes underwent positive selection in chimpanzee evolution than in human evolution. PNAS vol. 104:7489-7494.

Bamshad M, and Wooding SP (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4:99-111.

Barnhart MK, Connor LM, and Marriott SJ (1997) Function of the human T-cell leukemia virus type 1 21-base-pair repeats in basal transcription. J Virol 71:337-344.

Barr CS, Newman TK, Becker ML, Parker CC, Champoux M, Lesch KP, Goldman D, Suomi SJ, and Higley JD (2003) The utility of the non-human primate; model for studying gene by environment interactions in behavioral research. Genes Brain Behav 2:336-340.

Barr CS, Newman TK, Shannon C, Parker C, Dvoskin RL, Becker ML, Schwandt M, Champoux M, Lesch KP, Goldman D, Suomi SJ, and Higley JD (2004) Rearing condition and rh5-HTTLPR interact to influence limbic-hypothalamic-pituitary-adrenal axis response to stress in infant macaques. Biol Psychiatry 55:733-738.

Barrett P, Ivanova E, Graham ES, Ross AW, Wilson D, Ple H, Mercer JG, Ebling FJ, Schuhler S, Dupre SM, Loudon A, and Morgan PJ (2006) Photoperiodic regulation of cellular retinoic acid-binding protein 1, GPR50 and nestin in tanycytes of the third ventricle ependymal layer of the Siberian hamster. J Endocrinol 191:687-698.

Barton RA (1996) Neocortex size and behavioural ecology in primates. Proc Biol Sci 263:173-177.

Barton RA (1998) Visual specialization and brain evolution in primates. Proc Biol Sci 265:1933-1937.

Battersby S, Ogilvie AD, Smith CA, Blackwood DH, Muir WJ, Quinn JP, Fink G, Goodwin GM, and Harmar AJ (1996) Structure of a variable number tandem repeat of the serotonin transporter gene and association with affective disorder. Psychiatr Genet 6:177-181.

Bell AC, West AG, and Felsenfeld G (1999) The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators. Cell 98:387-396.

Belting HG, Cooduvalli SS, and Ruddle FH (1998) Modification of expression and cis-regulation of Hoxc8 in the evolution of diverged axial morphology. Proc. Natl. Acad. Sci. USA 95:2355–2360.

Bennett AJ, Lesch KP, Heils A, Long JC, Lorenz JG, Shoaf SE, Champoux M, Suomi SJ, Linnoila MV, and Higley JD (2002) Early experience and serotonin transporter gene variation interact to influence primate CNS function. Mol Psychiatry 7:118-122.

Berg ES, and Olaisen B (1993) Characterization of the COL2A1 VNTR polymorphism. Genomics 16:350-354.

Bigger CB, Melnikova IN, and Gardner PD (1997) Sp1 and Sp3 Regulate Expression of the Neuronal Nicotinic Acetylcholine Receptor beta 4 Subunit Gene. J Biol Chem 272: 25976-25982.

Bird A (1986) CpG-rich islands and the function of DNA methylation. Nature 321:209-213.

Bock CB, Paulsen M, Tierling S, Mikeska T, Lengauer T, and Walter J (2006) CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats and predicted DNA structure. PLoS Genet 2:243-252.

Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, and Rubin EM (2003) Phylogenetic shadowing of primate sequences to find functional regions of the human genome. Science 299:1391-1394.

Bouwman P, and Philipsen S (2002) Regulation of the activity of Sp1-related transcription factors. Mol Cell Endocrinol 195:27-38.

Bowen S, Roberts C, and Wheals AE (2005) Patterns of polymorphism and divergence in stress-related yeast proteins. Yeast 22:659-668.

Boyer SH, Crosby EF, Noyes AN, Fuller GF, Leslie SE, Donaldson LJ, Vrablik GR, Schaefer EW, Jr., and Thurmon TF (1971) Primate hemoglobins: Some

sequences and some proposals concerning the character of evolution and mutation. Biochem Genet 5:405-448.

Brickman JM, Clements M, Tyrell R, McNay D, Woods K, Warner J, Stewart A, Beddington RSP, and Dattani M (2001) Molecular effects of novel mutations in Hesx1/HESX1 associated with human pituitary disorders. Development 128:5189-5199.

Burke TW, Willy PJ, Kutach AK, Butler JE, and Kadonaga JT (1998). The DPE, a conserved downstream core promoter element that is functionally analogous to the TATA box. Cold Spring Harb Symp Quant Biol 63:75-82.

Bush G, Vogt BA, Holmes J, Dale AM, Greve D, Jenike MA, and Rosen BR (2002) Dorsal anterior cingulate cortex: A role in reward-based decision-making. Proc Natl Acad Sci U S A 99:523-528.

Bussière T, and Hof PR (2004) The aging brain: morphomolecular senescence of cortical circuits. Trends in Neurosciences 27: 607-613.

Cáceres M, Lachuer J, Zapala MA, Redmond JC, Kudo LC, Geschwind DH, Lockhart DJ, Preuss TM, and Barlow C (2003) Elevated gene expression levels distinguish human from non-human primate brains. Proc Natl Acad Sci U S A 100: 13030-13035.

Carroll SB (2000) Endless Forms: The Evolution of Gene Regulation and Morphological Diversity. Minireview. Cell 101:577-580.

Carroll SB (2005) Evolution at two levels: on genes and form. PLoS Biol 3:e245.

Caspi A and Moffitt TE. (2006). Gene-environment interactions in psychiatry: joining forces with neuroscience. Nat Rev Neurosci. 7(7):583-590. Review.

Cerrone GE, Caputo M, Lopez AP, Gonzalez C, Massa C, Cedola N, Targovnik HM, and Frechtel GD (2004) Variable number of tandem repeats of the insulin gene determines susceptibility to latent autoimmune diabetes in adults. Mol Diagn 8:43-49.

Chen-Park FE, Huang DB, Noro B, Thanos D, and Ghosh G (2002) The kappa B DNA sequence from the HIV long terminal repeat functions as an allosteric regulator of HIV transcription. J Biol Chem 277:24701-24708.

Chen FC, Vallender EJ, Wang H, Tzeng CS, and Li WH (2001) Genomic Divergence Between Human and Chimpanzee Estimated from Large-Scale Alignments of Genomic Sequences. The Journal of Heredity 92:481-489.

Chernukhin I, Shamsuddin S, El-Kady AI, Paul A, Lobanenkov V, Klenova E, Carne AF, and Robinson AF (2000) Physical and functional interaction between two pluripotent proteins, the Y-box DNA/RNA-binding factor, YB-1, and the multivalent zinc finger factor, CTCF. J Biol Chem 275:29915-29921.

Chernukhin I, Shamsuddin S, Kang SY, Bergström R, Kwon YW, Yu W, Whitehead J, Mukhopadhyay R, Docquier F, Farrar D, Morrison I, Vigneron M, Wu SY, Chiang CM, Loukinov D, Lobanenkov V, Ohlsson R, Klenova E (2007) CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. Mol Cell Biol. 27(5):1631-1648 (Epub ahead of print, January).

Chomczynski P (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extract. Analytical biochemistry 162:156.

Clark AG, Glanowski S, Nielsen R, Thomas D, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferriera S, Wang G, Zheng X, White TJ, Sninsky JJ, Adams MD, and Cargill M (2003) Inferring Nonneutral Evolution from Human-Chimp-Mouse Orthologous Gene Trios. Science 302: 1960-1963.

Cohen HT, Bossone SA, Zhu G, McDonald GA, and Sukhatme VP (1997) Sp1 Is a Critical Regulator of the Wilms' tumor-1 Gene. J Biol Chem 272:2901-2913.

Collier-Baker E, Davis JM, Nielsen M, and Suddendorf T (2006) Do chimpanzees (*Pan troglodytes*) understand single invisible displacement? Anim Cogn 9:55-61.

Congdon E, Lesch KP, and Canli T (2007) Analysis of DRD4 and DAT polymorphisms and behavioral inhibition in healthy adults: Implications for impulsivity. Am J Med Genet B Neuropsychiatr Genet (Epub ahead of print, May)

Conneally PM (1994) Human genetic polymorphisms. Dev Biol Stand 83:107-110.

Conrad SE, and Botchan MR (1982) Isolation and characterization of human DNA fragments with nucleotide sequence homologies with the simian virus 40 regulatory region. Mol Cell Biol 2:949-965.

Cook EH, Jr., Stein MA, Krasowski MD, Cox NJ, Olkon DM, Kieffer JE, and Leventhal BL (1995) Association of attention-deficit disorder and the dopamine transporter gene. Am J Hum Genet 56:993-998.

Cress WD, and Seto E (2000) Histone deacetylases, transcriptional control, and cancer. J Cell Physiol 184:1-16.

Csordas A (1990) On the biological role of histone acetylation. Biochem. J. 265:23-38.

Curran S, Purcell S, Craig I, Asherson P, and Sham P (2005) The serotonin transporter gene as a QTL for ADHD. Am J Med Genet B Neuropsychiatr Genet 134:42-47.

Dean A. (2006) On a chromosome far, far away: LCRs and gene expression. Trends in Genetics 22 (1): 38:45

D'Souza U, and Craig IW (2006) Functional Polymorphisms in Dopamine and Serotonin Pathway Genes. Human Mutation 27:1-13.

D'Souza UM, Russ C, Tahir E, Mill J, McGuffin P, Asherson PJ, and Craig IW (2004) Functional effects of a tandem duplication polymorphism in the 5'flanking region of the DRD4 gene. Biol Psychiatry 56:691-697.

Dainton M, and Macho GA (1999) Did knuckle walking evolve twice? J Hum Evol 36:171-194.

Davidson S, Miller KA, Dowell A, Gildea A, and MacKenzie A (2006) A remote and highly conserved enhancer supports amygdala specific expression of the gene encoding the anxiogenic neuropeptide substance-P. Mol Psychiatry 11: 410-421.

Demiralp T, Herrmann CS, Erdal ME, Ergenoglu T, Keskin YH, Ergen M, and Beydagi H (2007) DRD4 and DAT1 polymorphisms modulate human gamma band responses. Cereb Cortex 17:1007-1019.

Dermitzakis ET, and Clark AG (2002) Evolution of Transcription Factor Binding Sites Regulatory Regions: Conservation and Turnover. Mol. Biol. Evol. 19:1114-1121.

Ding YC, Chi HC, Grady DL, Morishima A, Kidd JR, Kidd KK, Flodman P, Spence MA, Schuck S, Swanson JM, Zhang YP, and Moyzis RK (2002) Evidence of positive selection acting at the human dopamine receptor D4 gene locus. Proc Natl Acad Sci U S A 99:309-314.

Dorus S, Vallender EJ, Evans PD, Anderson JR, Gilbert SL, Mahowald M, Wyckoff GJ, Malcom CM, and Lahn BT (2004) Accelerated evolution of nervous system genes in the origin of *Homo sapiens*. Cell 119:1027-1040.

Dunbar RIM (2003) Psychology. Evolution of the social brain. Science 302:1160-1161.

Dunbar RIM (1992) Neocortex Size as a Constraint on Group-Size in Primates. J Hum Evol 22:469-493.

Dunbar RIM, and Bever J (1998) Neocortex size predicts group size in carnivores and some insectivores. Ethology 104:695-708.

Dutrillaux B (1980) Chromosomal evolution of the great apes and man. J Reprod Fertil (Suppl) 28:105-111.

Eberharter A, and Becker PB (2002) Histone acetylation: a switch between repressive and permissive chromatin. EMBO Rep 3:224-229.

Eley TC, Sugden K, Corsico A, Gregory AM, Sham P, McGuffin P, Plomin R, Craig IW. (2004).Gene–environment interaction analysis of serotonin system markers with adolescent depression. Molecular Psychiatry 9(10):908-915

Enard W, Fassbender A, Model F, Adorján P, Pääbo S, and Olek A (2004) Differences in DNA methylation patterns between humans and chimpanzees. Current Biology 14:R148-R149.

Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, and Paabo S (2002a) Intra- and interspecific variation in primate gene expression patterns. Science 296:340-343.

Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, and Paabo S (2002b) Molecular evolution of FOXP2, a gene involved in speech and language. Nature 418:869-872.

Erblich J, Lerman C, Self DW, Diaz GA, and Bovbjerg DH (2004) Stress-induced cigarette craving: effects of the DRD2 TaqI RFLP and SLC6A3 VNTR polymorphisms. Pharmacogenomics J 4:102-109.

Escalante AA, Barrio E, and Ayala FJ (1995) Evolutionary origin of human and primate malarias: evidence from the circumsporozoite protein gene. Mol Biol Evol 12:616-626.

Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5:164-166.

Filippova G, Lindblom A, Meincke L, Klenova E, Neiman P, Collins S, Doggett N, and Lobanenkov V (1998) A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. Genes Chromosomes Cancer 22:26-36.

Fiskerstrand CE, Lovejoy EA, and Quinn JP (1999) An intronic polymorphic domain often associated with susceptibility to affective disorders has allele dependent differential enhancer activity in embryonic stem cells. FEBS Lett. 458:171-174.

Foley RA, and Lee PC (1991) Ecology and energetics of encephalization in hominid evolution. Philos Trans R Soc Lond B Biol Sci 334:223-231.

Fullerton SM, Bartoszewicz A, Ybazeta G, Horikawa Y Bell GI, Kidd KK, Cox NJ, Hudson RR and Di Rienzo A. 2002. Geographic and haplotype structure of candidate type 2 diabetes-susceptibility variants at the calpain-10 locus. Am. J. Hum. Genet. 70:1096–1106.

Gagneux P (2002) The genus *Pan*: population genetics of an endangered outgroup Trends in Genetics 18(7):327-330.

Galeeva AR, Gareeva AE, Iur'ev EB, and Khusnutdinova EK (2002) VNTR polymorphisms of the serotonin transporter and dopamine transporter genes in male opiate addicts. Mol Biol (Mosk) 36:593-598.

Galvani AP, and Novembre J (2005) The evolutionary history of the CCR5-Delta 32 HIV-resistance mutation. Microbes and Infection 7:302-309.

Garver FA, and Talmage DW (1975) Comparative immunochemical studies of primate hemoglobins. Biochem Genet 13:743-757.

Gelernter J, Cubells JF, Kidd JR, Pakstis AJ, and Kidd KK (1999) Population studies of polymorphisms of the serotonin transporter protein gene. Am J Med Genet 88:61-66.

Gelernter J, Kranzler HR, Satel SL, and Rao PA (1994) Genetic association between dopamine transporter protein alleles and cocaine-induced paranoia. Neuropsychopharmacology 11:195-200.

Gerrard L, Howard M, Paterson T, Thippeswamy T, Quinn JP, and Haddley K (2005) A proximal E-box modulates NGF effects on rat PPT-A promoter activity in cultured dorsal root ganglia neurones. Neuropeptides 39:475-483.

Gervais M, and Wilson DS (2005) The evolution and functions of laughter and humor: a synthetic approach. Q Rev Biol 80:395-430.

Geyer PK, and Corces VG (1987) Separate regulatory elements are responsible for the complex pattern of tissue-specific and developmental transcription of the yellow locus in *Drosophila melanogaster*. Genes Dev 1:996-1004.

Gillies SD, Morrison SL, Oi VT, and Tonegawaa S (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. Cell 33:717-728.

Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, and Heintz N (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. Nature 425: 917-925.

Goodman M (1967) Effects of Evolution on Primates Macromolecules. Primates 8:1-22.

Goodman M, Barnabas J, Matsuda G, and Moore JW (1971) Molecular Evolution in the Descent of Man. Nature Biotechnology 233:604 - 613.

Gornick MC, Addington A, Shaw P, Bobb AJ, Sharp W, Greenstein D, Arepalli S, Castellanos FX, and Rapoport JL (2007) Association of the dopamine receptor D4 (DRD4) gene 7-repeat allele with children with attention-deficit/hyperactivity disorder (ADHD): an update. Am J Med Genet B Neuropsychiatr Genet 144:379-382.

Greenwood TA, and Kelsoe JR (2003) Promoter and intronic variants affect the transcriptional regulation of the human dopamine transporter gene. Genomics 82: 511-519.

Guindalini C, Howard M, Haddley K, Laranjeira R, Collier D, Ammar N, Craig I, O'Gara C, Bubb VJ, Greenwood T, Kelsoe J, Asherson P, Murray RM, Castelo A, Quinn JP, Vallada H, and Breen G (2006) A dopamine transporter gene functional variant associated with cocaine abuse in a Brazilian sample. Proc Natl Acad Sci U S A 103:4552-4557.

Gunnell GF, and Miller ER (2001) Origin of Anthropoidea: dental evidence and recognition of early anthropoids in the fossil record, with comments on the Asian anthropoid radiation. Am. J. Phys. Anthropol. 114:177-191.

Haddley K, Vasiliou AS, Ali FR, Paredes UM, Bubb VJ, and Quinn JP (2007) Molecular Genetics of Monoamine Transporters: Relevance to Brain Disorders. Neurochem Res (Epub ahead of print, October).

Harding RM, Boyce AJ, and Clegg JB (1992) The Evolution of Tandemly Repetitive DNA: Recombination Rules. Genetics 132:847-859.

Hariri AR, and Holmes A (2006) Genetics of emotional regulation: the role of the serotonin transporter in neural function. Trends Cogn Sci *10*:182-191.

Hariri AR, Mattay VS, Tessitore A, Kolachana B, Fera F, Goldman D, Egan MF, and Weinberger DR (2002a) Serotonin transporter genetic variation and the response of the human amygdala. Science 297:400-403.

Hariri AR, Tessitore A, Mattay VS, Fera F, and Weinberger DR (2002b) The amygdala response to emotional stimuli: a comparison of faces and scenes. Neuroimage 17:317-323.

Harris EE (2000) Molecular systematics of the Old World monkey tribe Papionini: analysis of the total available genetic sequences. J Hum Evol 38:235-256.

He X, and Rosenfeld MG (1991) Mechanisms of complex transcriptional regulation: implications for brain development. Review. Neuron 7:183-196.

Healy SD, and Hurly TA (2004) Spatial learning and memory in birds. Brain Behav. Evol. 63:211–220.

Heils A, Mobner R, and Lesch KP (1997) The human serotonin transporter gene polymorphism: basic research and clinical implications. J Neural Transm 104:1005-1014.

Heils A, Teufel A, Petri S, Seemann M, Bengel D, Bailing U, Riederer P, and Lesch KP (1995) Functional promoter and polyadenylation site mapping of the human serotonin (5-HT) transporter gene. J Neural Transm [Gen Sect] 102:247-254.

Heils A, Teufel A, Petri S, Stöber G, Riederer P, and Lesch KP (1996) Allelic Variation of Human Serotonin Transporter Gene Expression. Journal of Neurochemistry 66:2621-2624.

Heils A, Wichems C, Mossner R, Petri S, Glatz K, Bengel D, Murphy DL, and Lesch KP (1998) Functional Characterization of the Murine Serotonin Transporter Gene Promoter in Serotonergic Raphe Neurons. J. Neurochem. 70:932-939.

Heinz A, Goldman D, Jones DW, Palmour R, Hommer D, Gorey JG, Lee KS, Linnoila M, and Weinberger DR (2000) Genotype influences in vivo dopamine transporter availability in human striatum. Neuropsychopharmacology 22:133-139.

Heissig F, Krause J, Bryk J, Khaitovich P, Enard W, and Paabo S (2005) Functional analysis of human and chimpanzee promoters. Genome Biol *6*:R57.

Herlenius E, and Lagercrantz H (2001) Neurotransmitters and neuromodulators during early human development. Early Hum Dev 65:21-37.

Hof PR, Nimchinsky EA, Perl DP, and Erwin JM (2001) An unusual population of pyramidal neurons in the anterior cingulate cortex of hominids contains the calcium-binding protein calretinin. Neurosci Lett 307:139–142.

Hood DB, Nayyar T, Ramesh A, Greenwood M, Inyang F (2000) Modulation in the developmental expression profile of Sp1 subsequent to transplacental exposure of fetal rats to desorbed benzo (a)pyrene following maternal inhalation. Inhal Toxicol. 12(6):511-535.

Houdart R (2005) Memory: a view of the history of central nervous system. Encephale 31:317-322.

Hranilovic D, Stefulj J, Schwab S, Borrmann-Hassenbach M, Albus M, Jernej B, and Wildenauer D (2004) Serotonin transporter promoter and intron 2 polymorphisms: relationship between allelic variants and gene expression. Biol Psychiatry 55:1090-1094.

Hull LA (1982) Modifications of chromatin structure in control of gene expression. Med Hypotheses 8:113-133.

Inoue-Murayama M, Adachi S, Mishima N, Mitani H, Takenaka O, Terao K, Hayasaka I, Ito S, and Murayama Y (2002a) Variation of variable number of tandem repeat sequences in the 3'-untranslated region of primate dopamine transporter genes that affects reporter gene expression. Neurosci Lett 334:206-210.

Inoue-Murayama M, Matsuura N, Murayama Y, Tsubota T, Iwasaki T, Kitagawa H, and Ito S (2002b) Sequence comparison of the dopamine receptor D4 exon III repetitive region in several species of the order Carnivora. J Vet Med Sci 64:747-749.

Inoue-Murayama M, Mishima N, Hayasaka I, Ito S, and Murayama Y (2006) Divergence of ape and human monoamine oxidase A gene promoters: comparative analysis of polymorphisms, tandem repeat structures and transcriptional activities on reporter gene expression. Neurosci Lett 405:207-11.

Inoue-Murayama M, Murayama Y, and Takenaka O (1998) Origin and divergence of tandem repeats of primate D4 dopamine receptor genes. Primates 39:217-224.

Jaenisch R, and Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Review. Nat Genet 33 Suppl:245-254.

Jeffreys AJ (1997) Spontaneous and induced minisatellite instability in the human genome. Clin Sci (Lond) 93:383-90.

Joffe TH, and Dunbar RI (1997) Visual and socio-cognitive information processing in primate brain evolution. Proc Biol Sci 264:1303-7.

Jones KE, and MacLarnon AM (2004) Affording larger brains: testing hypotheses of mammalian brain evolution on bats. Am Nat 164:E20-31.

Kamakura S, Iwaki A, Matsumoto M, Fukumaki Y. (1997) Cloning and characterisation of the 5' flanking region of the human dopamine D4 receptor gene. Biochem Biophys Res Commun 235:321–326.

Kappelman J (1996) The evolution of body mass and relative brain size in fossil hominids. J Hum Evol 30: 243–276.

Kent WJ (2002) BLAT--the BLAST-like alignment tool. Genome Res 12:656-664.

Khaitovich P, Muetzel B, She X, Lachmann M, Hellmann I, Dietzsch J, Steigele S, Do HH, Weiss G, Enard W, Heissig F, Arendt T, Nieselt-Struwe K, Eichler EE, and Paabo S (2004b) Regional patterns of gene expression in human and chimpanzee brains. Genome Res 14:1462-1473.

Khaitovich P, Enard W, Lachmann M, and Pääbo S. 2006. Evolution of primate gene expression. Nature Reviews genetics 7: 693-702.

Kim SJ, Badner J, Cheon KA, Kim BN, Yoo HJ, Kim SJ, Cook E, Jr., Leventhal BL, and Kim YS (2005) Family-based association study of the serotonin transporter gene polymorphisms in Korean ADHD trios. Am J Med Genet B Neuropsychiatr Genet 139:14-18.

King MC, and Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188:107-116.

Klenova E, Scott AC, Roberts J, Shamsuddin S, Lovejoy EA, Bergmann S, Bubb VJ, Royer HD, and Quinn JP (2004) YB-1 and CTCF differentially regulate the 5-HTT polymorphic intron 2 enhancer which predisposes to a variety of neurological disorders. J Neurosci 24:5966-5973.

Klug A, Rhodes D, Smith J, Finch JT, and Thomas JO (1980) A low resolution structure for the histone core of the nucleosome. Nature 287:509-516.

Kohno K, Izumi H, Uchiumi T, Ashizuka M, and Kuwano M (2003) The pleiotropic functions of the Y-box-binding protein, YB-1. BioEssays 25:691-698.

Kornberg RD, and Klug A (1981) The nucleosome. Sci Am 244:52-64.

Kouprina N, Pavlicek A, Mochida GH, Solomon G, Gersch W, Yoon YH, Collura R, Ruvolo M, Barrett JC, Woods CG, Walsh CA, Jurka J, and Larionov V (2004) Accelerated evolution of the ASPM gene controlling brain size begins prior to human brain expansion. PLoS Biol 2:E126.

Kremer I, Bachner-Melman R, Reshef A, Broude L, Nemanov L, Gritsenko I, Heresco-Levy U, Elizur Y, and Ebstein RP (2005) Association of the serotonin transporter gene with smoking behavior. Am J Psychiatry 162:924-930.

Krezel W, Ghyselinck N, Samad TA, Dupe V, Kastner P, Borrelli E, and Chambon P (1998) Impaired locomotion and dopamine signaling in retinoid receptor mutant mice. Science 279:863-867.

Kurokawa D, Sakurai Y, Inoue A, Nakayama R, Takasaki N, Suda Y, Miyake T, Amemiya C T and Aizawa S (2006) Evolutionary constraint on Otx2 neuroectoderm enhancers-deep conservation from skate to mouse and unique divergence in teleost. Proc Natl Acad Sci U S A 51: 19350–19355

Lakatos K, Nemoda Z, Toth I, Ronai Z, Ney K, Sasvari-Szekely M, and Gervai J (2002) Further evidence for the role of the dopamine D4 receptor (DRD4) gene in attachment disorganization: interaction of the exon III 48-bp repeat and the -521 C/T promoter polymorphisms. Mol Psychiatry 7:27-31.

Larsen SA, Mogensen L, Dietz R, Baagoe HJ, Andersen M, Werge T, and Rasmussen HB (2005) Identification and characterization of tandem repeats in exon III of dopamine receptor D4 (DRD4) genes from different mammalian species. DNA Cell Biol 24:795-804.

Lasky-Su JA, Faraone SV, Glatt SJ, and Tsuang MT (2005) Meta-analysis of the association between two polymorphisms in the serotonin transporter gene and affective disorders. Am J Med Genet B Neuropsychiatr Genet 133:110-115.

Latchman DS (1992) Gene regulation. BMJ 304:1103-1105.

Latchman DS (1993) Transcription factors: an overview. Int J Exp Pathol 74:417-422.

Latchman DS (1996) Transcription-factor mutations and disease. N Engl J Med 334:28-33.

Lesch KP, Balling U, Gross J, Strauss K, Wolozin BL, Murphy DL, and Riederer P (1994) Organization of the human serotonin transporter gene. J Neural Transm Gen Sect 95:157-162.

Lesch KP, Meyer J, Glatz K, Flugge G, Hinney A, Hebebrand J, Klauck SM, Poustka A, Poustka F, Bengel D, Mossner R, Riederer P, and Heils A (1997) The 5-HT transporter gene-linked polymorphic region (5-HTTLPR) in evolutionary perspective: alternative biallelic variation in rhesus monkeys. Rapid communication. J Neural Transm 104:1259-1266.

Livak KJ, Rogers J, and Lichter JB (1995) Variability of dopamine D4 receptor (DRD4) gene sequence within and among nonhuman primate species. Proc Natl Acad Sci U S A 92:427-431.

Lovejoy EA, Scott AC, Fiskerstrand CE, Bubb VJ, and Quinn JP (2003) The serotonin transporter intronic VNTR enhancer correlated with a predisposition to affective disorders has distinct regulatory elements within the domain based on the primary DNA sequence of the repeat unit. Eur J Neurosci 17:417-420.

Lobanenkov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, Goodwin GH (1990) A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5'-flanking sequence of the chicken c-myc gene. Oncogene. 5(12):1743-1753.

Luo ZX, Ji Q, Wible JR, and Yuan CX (2003) An Early Cretaceous Tribosphenic Mammal and Metatherian Evolution Science 302:1934 - 1940

Lusser A (2002) Acetylated, methylated, remodeled: chromatin states for gene regulation. Curr Opin Plant Biol 5:437-443.

Lynch M, and Conery JS (2000) The Evolutionary Fate and Consequences of Duplicate Genes. Science 290: 1151-1155.

Mabandla MV, Dobson B, Johnson S, Kellaway LA, Daniels WM, and Russell VA (2007) Development of a mild prenatal stress rat model to study long term effects on neural function and survival. Metab Brain Dis (Epub ahead of print, August).

MacKenzie A, and Quinn JP (1999) A serotonin transporter gene intron 2 polymorphic region, correlated with affective disorders, has allele-dependent differential enhancer-like properties in the mouse embryo. Proc Natl Acad Sci U S A 96:15251–15255.

MacKenzie A, and Quinn JP (2004) Post-genomic approaches to exploring neuropeptide gene mis-expression in disease. Neuropeptides 38:1-15.

Mage RG, Newman BA, Harindranath N, Bernstein KE, Becker RS, and Knight KL (1989) Evolutionary conservation of splice sites in sterile C mu transcripts and of immunoglobulin heavy chain (IgH) enhancer region sequences. Mol Immunol 26:1007-1010.

Mahncke HW, Bronstone A, and Merzenich MM (2006) Brain plasticity and functional losses in the aged: scientific bases for a novel intervention. Prog Brain Res 157:81-109.

Manson JH, and Wrangham RW (1991) Intergroup Aggression in Chimpanzees and Humans. Current anthropology 32: 369-390.

Matsuzawa T (2007) Comparative cognitive development. Developmental Science 10: 97-103.

McCracken JT, Smalley SL, McGough JJ, Crawford L, Del'Homme M, Cantor RM, Liu A, and Nelson SF (2000) Evidence for linkage of a tandem duplication polymorphism upstream of the dopamine D4 receptor gene (DRD4) with attention deficit hyperactivity disorder (ADHD). Mol Psychiatry 5:531-536.

McGinnis W, Jack T, Chadwick R, Regulski M, and Bergson C (1990) Establishment and maintenance of position-specific expression of the Drosophila homeotic selector gene Deformed. Adv Genet 27:363-402.

McQueen JK, Wilson H, Sumner BEH, and Fink G (1999) Serotonin transporter (SERT)mRNA and binding site densities in male rat brain affected by sex steroids. Mol Brain Res 63:241-247.

Meloni R, Albanese V, Ravassard P, Treilhou F, Mallet J. (1998). A rare allele of a tetranucleotide polymorphic microsatellite, located in the first intron of the tyrosine hydroxylase gene, is associated with schizophrenia and shows transcription regulatory activity *in vitro*. Schizophrenia Research 29(1): 129-129.

Mertens P, Alfonso-Jaume A, Steinman K, and Lovett DH (1998) A synargistic interaction of transcription factors AP2 and YB-1 regulates gelatinase A enhancer-depedent transcription. The Journal of Biological Chemestry 273:32957-32965.

Meyer A, and Zardoya R (2003) Recent advances in the (molecular) phylogeny of vertebrates. Annual Review of Ecology, Evolution, and Systematics 34:311-338.

Michelhaugh SK, Fiskerstrand C, Lovejoy E, Bannon MJ, and Quinn JP (2001) The dopamine transporter gene (SLC6A3) variable number of tandem repeats domain enhances transcription in dopamine neurons. J Neurochem 79:1033-1038.

Mill J, Asherson P, Browes C, D'Souza U, and Craig I (2002) Expression of the dopamine transporter gene is regulated by the 3' UTR VNTR: Evidence from brain and lymphocytes using quantitative RT-PCR. Am J Med Genet 114:975-979.

Miller GM, Bendor J, Tiefenbacher S, Yang H, Novak MA, and Madras BK (2004) A mu-opioid receptor single nucleotide polymorphism in rhesus monkey: association with stress response and aggression. Mol Psychiatry 9:99-108.

Miller GM, De La Garza R, 2nd, Novak MA, and Madras BK (2001) Single nucleotide polymorphisms distinguish multiple dopamine transporter alleles in primates: implications for association with attention deficit hyperactivity disorder and other neuropsychiatric disorders. Mol Psychiatry 6:50-58.

Miller GM and Madras BK (2002) Polymorphisms in the 3'-untranslated region of human and monkey dopamine transporter genes affect reporter gene expression. Molecular Psychiatry 7: 44–55.

Michelhaugh SK, Fiskerstrand C, Lovejoy E, Bannon MJ and Quinn JP (2001) The dopamine transporter gene (SLC6A3) variable number of tandem repeats domain enhances transcription in dopamine neurons. J Neurochem 79:1033-1038.

Mogensen L, Kinze CC, Werge T, and Rasmussen HB (2006) Identification and characterization of a tandem repeat in exon III of the dopamine receptor D4 (DRD4) gene in cetaceans. J Hered 97:279-284.

Momozawa Y, Takeuchi Y, Kusunose R, Kikusui T, and Mori Y (2005) Association between equine temperament and polymorphisms in dopamine D4 receptor gene. Mamm Genome 16:538-544.

Morrison JH, and Hof PR (1997) Life and Death of Neurons in the Aging Brain. Science 278:412 - 419.

Mulder EJ, Anderson GM, Kema IP, Brugman AM, Ketelaars CE, de Bildt A, van Lang ND, den Boer JA, and Minderaa RB (2005) Serotonin transporter intron 2 polymorphism associated with rigid-compulsive behaviors in Dutch individuals with pervasive developmental disorder. Am J Med Genet B Neuropsychiatr Genet 133:93-96.

Murgatroyd C, Wigger A, Frank E, Singewald N, Bunck M, Holsboer F, Landgraf R, and Spengler D (2004) Impaired repression at a vasopressin promoter polymorphism underlies overexpression of vasopressin in a rat model of trait anxiety. J Neurosci 24:7762-7770.

Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryderk OA and O'Brien SJ. (2001). Molecular phylogenetics and the origins of placental mammals. Nature 409(6820):614-618.

Myers RL, Airey DC, Manier DH, Shelton RC, and Sanders-Bush E (2007) Polymorphisms in the Regulatory Region of the Human Serotonin 5-HT2A Receptor Gene (HTR2A) Influence Gene Expression. Biol Psychiatry 61:167-173.

Nakamura Y, Koyama K, and Matsushima M (1998) VNTR (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators. J Hum Genet 43:149-52.

Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, and Bird A (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. Nature 393:386-389.

Ng HH, and Bird A (1999) DNA methylation and chromatin modification. Curr Opin Genet Dev 9:158-163.

Nie Z, Chen S, Kumar R, and Zack DJ (1996) RER, an evolutionarily conserved sequence upstream of the rhodopsin gene, has enhancer activity. J Biol Chem 271:2667-2675.

Nimchinsky E, Gilissen E, Allman JM, Perl DP, Erwin JM, and Hof PR (1999) A neuronal morphologic type unique to humans and great apes. Proc Nat Acad Sci 96:5268-5273.

Nishizaki Y, Shimazu K, Kondoh H, and Sasaki H (2001) Identification of essential sequence motifs in the node/notochord enhancer of Foxa2 (Hnf3beta) gene that are conserved across vertebrate species. Mech Dev 102:57-66.

Norman JT, Lindahl GE, Shakib K, En-Nia A, Yilmaz E, and Mertens PR (2001) The Y-box binding protein YB-1 suppresses collagen a 1(I) gene transcription via an evolutionarily conserved regulatory element in the proximal promoter. J Biol Chem 276:29880-29890.

Novina CO, and Ananda LR (1996) Core promoters and transcriptional control. Review. Trends in Genetics 12:351-355.

Oak JN, Oldenhof J, and Van Tol HH (2000) The dopamine D(4) receptor: one decade of research. Eur J Pharmacol 405:303-327.

Ohara K, Suzuki Y, Ochiai M, Tsukamoto T, Tani K, and Ohara K (1999) A variable-number-tandem-repeat of the serotonin transporter gene and anxiety disorders. Prog Neuropsychopharmacol Biol Psychiatry 23:55-65.

Ohlsson R, Renkawitz R, and Lobanenkov V (2001) CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends in Genetics 17 (9): 520-527.

Ohno S (1972) Simplicity of mammalian regulatory systems. Dev Biol. 27:131-136.

Okamoto S, Sherman K, Bai G, and Lipton SA (2002) Effect of the ubiquitous transcription factors, SP1 and MAZ, on NMDA receptor subunit type 1 (NR1) expression during neuronal differentiation. Brain Res Mol Brain Res 107:89-96.

Olson MV (1999) When less is more: gene loss as an engine of evolutionary change. Am J Hum Genet 64:18-23.

Olson MV, and Varki A (2003) Sequencing the chimpanzee genome: insights into human evolution and disease. Nat Rev Genet 4:20-28.

Olson MV, and Varki A (2004) Genomics. The chimpanzee genome--a bittersweet celebration. Science 305:191-2.

Ooi L, and Wood IC (2007) Chromatin crosstalk in development and disease: lessons from REST. Review. Nat Rev Genet 8:544-554.

Orphanides G, Lagrange T, and Reinberg D (1996) The general transcription factors of RNA polymerase II. Genes Dev 10:2657-83.

Page SL, Goodman M. 2001. Catarrhine phylogeny: noncoding DNA evidence for a diphyletic origin of the mangabeys and for a human-chimpanzee clade. Mol Phylogenet Evol. 8(1):14-25.

Palmour RM, Cronin JE, Childs A, and Grunbaum BW (1980) Studies of primate protein variation and evolution: microelectrophoretic detection. Biochem Genet 18:793-808.

Park DH, Kim CS, Kim YS, Sung BJ, Lee JG, Chung DS, and Lee SE (2004) A Polymorphism in the Serotonin-Transporter Gene and Elite Endurance Athlete Status. Medicine & Science in Sports & Exercise 36:S40.

Parks CL, and Shenk T (1996) The serotonin 1a receptor gene contains a TATA-less promoter that responds to MAZ and Sp1. J Biol Chem 271:4417-4430.

Parvin JD, and Sharp PA (1993) DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. Cell 73:533-540.

Payton A, Gibbons L, Davidson Y, Ollier W, Rabbitt P, Worthington J, Pickles A, Pendleton N, and Horan M (2005) Influence of serotonin transporter gene polymorphisms on cognitive decline and cognitive abilities in a nondemented elderly population. Mol Psychiatry 10:1133-1139.

Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, and Rubin EM (2006) *In vivo* enhancer analysis of human conserved non-coding sequences. Letters. Nature 444(7118):499-502.

Pillai RS, Bhattacharyya SN, Filipowicz W. 2007. Repression of protein synthesis by miRNAs: how many mechanisms?Trends Cell Biol. 17(3):118-126.

Posner MI, and Rothbart MK (1998) Attention, self-regulation and consciousness. Philos Trans R Soc Lond B Biol Sci 353:1915-1927.

Poux C, and Douzery EJ (2004) Primate phylogeny, evolutionary rate variations, and divergence times: a contribution from the nuclear gene IRBP. Am J Phys Anthropol 124:1-16.

Prabhakar S, Noonan JP, Paabo S, and Rubin EM (2006a) Accelerated evolution of conserved noncoding sequences in humans. Science 314:786.

Prabhakar S, Poulin F, Shoukry M, Afzal V, Rubin EM, Couronne O, and Pennacchio LA (2006b) Close sequence comparisons are sufficient to identify human cis-regulatory elements. Genome Res 16:855-863.

Preuss TM, Caceres M, Oldham MC, and Geschwind DH (2004) Human brain evolution: insights from microarrays. Nat Rev Genet 5:850-860.

Pugh BF, and Tjian R (1991) Transcription from a TATA-less promoter requires a multisubunit TFIID complex. Genes Dev 5:1935-1945.

Pugliese A, Zeller M, Fernandez A, Jr., Zalcberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, and Patel DD (1997) The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. Nat Genet 15:293-297.

Quinn JP (1996) Neuronal-specific gene expression-the interaction of both positive and negative transcriptional regulators. Prog Neurobiol 50:363-379.

Raghanti MA, Stimpson CD, Marcinkiewicz JL, Erwin JM, Hof PR, and Sherwood CC (2007) Differences in Cortical Serotonergic Innervation among Humans, Chimpanzees, and Macaque Monkeys: A Comparative Study. Cereb Cortex. (Epub ahead of print, June)

Reader SM, and Laland KN (2002) Social intelligence, innovation, and enhanced brain size in primates. Proc Natl Acad Sci U S A 99:4436-4441.

Reif A, and Lesch KP (2003) Toward a molecular architecture of personality. Behav Brain Res 139:1-20.

Riegler M, Sidhu M, Miller WJ, and O'Neill SL (2005) Evidence for a Global Wolbachia Replacement in Drosophila melanogaster. Current Biology Vol. 15:1428-1433.

Rightmire GP (2004) Brain size and encephalization in early to Mid-Pleistocene *Homo*. Am J Phys Anthropol 124:109-23.

Roberts J, Scott AC, Howard MR, Breen G, Bubb VJ, Klenova E, and Quinn JP (2007) Differential regulation of the serotonin transporter gene by lithium is mediated by transcription factors, CCCTC binding protein and Y-box binding protein 1, through the polymorphic intron 2 variable number tandem repeat. J Neurosci 27:2793-2801.

Rockman MV, Hahn MW, Soranzo N, Zimprich F, Goldstein DB, and Wray GA (2005) Ancient and recent positive selection transformed opioid cis-regulation in humans. PLoS Biol 3:e387.

Rockman MV, and Wray GA (2002) Abundant raw material for cis-regulatory evolution in humans. Mol. Biol. Evol. 19:1991-2004.

Ronai Z, Guttman A, Keszler G, and Sasvari-Szekely M (2004) Capillary electrophoresis study on DNA-protein complex formation in the polymorphic 5' upstream region of the dopamine D4 receptor (DRD4) gene. Curr Med Chem 11:1023-1029.

Rosenberg K, and Trevathan W (2002) Birth, obstetrics and human evolution. Bjog 109:1199-1206.

Roth G, and Dicke U (2005) Evolution of the brain and intelligence. Trends Cogn Sci 9:250-257.

Sabatini MJ, Ebert P, Lewis DA, Levitt P, Cameron JL, and Mirnics K (2007) Amygdala gene expression correlates of social behavior in monkeys experiencing maternal separation. J Neurosci 27:3295-3304.

Sacchetti P, Mitchell TR, Granneman JG, and Bannon MJ (2001) Nurr1 enhances transcription of the human dopamine transporter gene through a novel mechanism. J Neurochem 76:1565-1572.

Sakai K, Nakamura M, Ueno S, Sano A, Sakai N, Shirai Y, and Saito N (2002) The silencer activity of the novel human serotonin transporter linked polymorphic regions. Neuroscience Letters 327: 13-16.

Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB, and Batzer MA (2003) Alu elements and hominid phylogenetics. Proc Natl Acad Sci U S A 100: 12787-12791.

Sander T, Harms H, Podschus J, Finckh U, Nickel B, Rolfs A, Rommelspacher H, and Schmidt LG (1997) Allelic association of a dopamine transporter gene polymorphism in alcohol dependence with withdrawal seizures or delirium. Biol Psychiatry 41:299-304.

Santini S, Boore JL, and Meyer A (2003) Evolutionary Conservation of Regulatory Elements in Vertebrate Hox Gene Clusters. Genome Research 13:1111-1122.

Sarich VM, and Wilson AC (1967) Rates of Albumin Evolution in Primates. Proc Natl Acad Sci U S A 58:142-148.

Sawaya BE, Khalili K, and Amini S (1998) Transcription of the human immunodeficiency virus type 1 (HIV-1) promoter in central nervous system

cells: effect of YB-1 on expression of the HIV-1 long terminal repeat. J Gen Virol 79:239-246.
Schoots O, and Van Tol HH (2003) The human dopamine D4 receptor repeat sequences modulate expression. Pharmacogenomics J 3:343-348.
Scott AF, Heath P, Trusko S, and Bayer SH (1984) The Sequence of the Gorilla Fetal Globin Genes: Evidence for Multiple Gene Conversions in Human Evolution. Mol. Biol. Evol. 1:371-389.
Seaman MI, Chang FM, Quinones AT, and Kidd KK (2000) Evolution of exon 1 of the dopamine D4 receptor (DRD4) gene in primates. J Exp Zool 288:32-38.
Seaman MI, Fisher JB, Chang F, and Kidd KK (1999) Tandem duplication polymorphism upstream of the dopamine D4 receptor gene (DRD4). Am J Med Genet 88:705-709.
Segal JA, Barnett JL, and Crawford DL (1999) Functional analysis of natural variation in Sp1 binding sites of a TATA-less promoter. J. Mol. Evol. 49:736-749.
Semendeferi K, Damasio H, Frank R, and Van Hoesen GW (1997) The evolution of the frontal lobes: a volumetric analysis based on three-dimensional reconstructions of magnetic resonance scans of human and ape brains. J Hum Evol 32:375-88.
Semendeferi K, Lu A, Schenker N, and Damasio H (2002) Humans and great apes share a large frontal cortex. Nat Neurosc. 5: 272-276.
Seto E, Lewis B, and Shenk T (1993) Interactions between transcription factors Sp1 and YY1. Nature 365:462-464.
Shanley DP, and Kirkwood TB (2001) Evolution of the human menopause. Bioessays 23:282-287.
Shashikant CS, Bolanowsky SA, Anand S, and Anderson SM (2007) Comparison of diverged Hoxc8 early enhancer activities reveals modification of regulatory interactions at conserved cis-acting elements. J Exp Zoolog B Mol Dev Evol 308:242-249.
Sherwood CC, Broadfield DC, Holloway RL, Gannon PJ, and Hof PR (2003) Variability of Broca's Area Homologue in African Great Apes: Implications for Language Evolution. Anat Rec 271(A):276-285.
Shilatifard A (2006) Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem 75:243-269.
Shimada MK, Inoue-Murayama M, Ueda Y, Maejima M, Murayama Y, Takenaka O, Hayasaka I, and Ito S (2004) Polymorphism in the second intron of dopamine receptor D4 gene in humans and apes. Biochem Biophys Res Commun 316:1186-1190.
Shimada S, Kitayama S, Walther D, and Uhl G (1992) Dopamine transporter mRNA: dense expression in ventral midbrain neurons. Brain Res Mol Brain Res 13:359-362.
Sisk CL, and Foster DL (2004) The neural basis of puberty and adolescence. Review. Nat Neurosci. 7:1040-1047.
Slightom JL, Chang LY, Koop BF, and Goodman M (1985) Chimpanzee fetal G gamma and A gamma globin gene nucleotide sequences provide further evidence of gene conversions in hominine evolution. Mol Biol Evol 2:370-389.
Smale ST (1997) Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. Biochim Biophys Acta 1351:73-88.

Smale ST, and Kadonaga JT (2003) The RNA polymerase II core promoter. Annu Rev Biochem 72:449-479.
Soeby K, Larsen SA, Olsen L, Rasmussen HB, and Werge T (2005) Serotonin transporter: evolution and impact of polymorphic transcriptional regulation. Am J Med Genet B Neuropsychiatr Genet 136:53-57.
Stanford CB (2006) The behavioral ecology of sympatric African apes: implications for understanding fossil hominoid ecology. Primates 47:91-101.
Stone JR, and Wray GA (2001) Rapid evolution of cis-regulatory sequences via local point mutations. Mol Biol Evol 18:1764-1770.
Stroud JC, Wu Y, Bates DL, Han A, Nowick K, Paabo S, Tong H, and Chen L (2006) Structure of the Forkhead Domain of FOXP2 Bound to DNA. Structure 14:159-166.
Sullivan B (1971) Structure, function and evolution of primate hemoglobins. II. A survey of the oxygen-binding properties. Comp Biochem Physiol B 40:359-80.
Suñé C, and García-Blanco MA (1995) Sp1 transcription factor is required for in vitro basal and Tat-activated transcription from the human immunodeficiency virus type 1 long terminal repeat. J Virol 69:6572-6576.
Suske G (1999) The Sp-family of transcription factors. Review. Gene. 238(2):291-300.
Szantai E, Szmola R, Sasvari-Szekely M, Guttman A, and Ronai Z (2005) The polymorphic nature of the human dopamine D4 receptor gene: a comparative analysis of known variants and a novel 27 bp deletion in the promoter region. BMC genetics 6:39.
Talianidis I, Tambakaki A, Toursounova J, and Zannis VI (1995) Complex Interactions between SPl Bound to Multiple Distal Regulatory Sites and HNF-4 Bound to the Proximal Promoter Lead to Transcriptional Activation of Liver-Specific Human APOCIII Gene? Biochemistry 34:10298-10309.
Thorpe SK, Holder RL, and Crompton RH (2007) Origin of human bipedalism as an adaptation for locomotion on flexible branches. Science 316:1328-1331.
Travers AA, and Klug A (1987) The bending of DNA in nucleosomes and its wider implications. Philos Trans R Soc Lond B Biol Sci 317:537-561.
Trefilov A, Berard J, Krawczak M, and Schmidtke J (2000) Natal dispersal in rhesus macaques is related to serotonin transporter gene promoter variation. Behav Genet 30:295-301.
Turek-Plewa J, and Jagodzinski PP (2005) The role of mammalian DNA methyltransferases in the regulation of gene expression. Cell Mol Biol Lett 10:631-647.
Turleau C, and de Grouchy J (1972) Karyotypes of man and the chimpanzee. Comparison of the topography of bands. Possible evolutive mechanisms. C R Acad Sci Hebd Seances Acad Sci D. 274:2355-2357.
Turner BM (1991) Histone acetylation and control of gene expression. J Cell Sci 99 (Pt 1):13-20.
Uchiumi T, Kohno K, Tanimura H, Matsuo K, Sato S, Uchida Y, and Kuwano M (1993) Enhanced expression of the human multidrug resistance 1 gene in response to UV light irradiation. Cell Growth Differ 4:147-157.
Van Tol HH (1998) Structural and functional characteristics of the dopamine D4 receptor. Adv Pharmacol 42:486-490.

VanNess SH, Owens MJ and Kilts CD (2005) The variable number of tandem repeats element in DAT1 regulates *in vitro* dopamine transporter density. BMC Genet. 27;6:55.

Vandenbergh DJ, Persico AM, Hawkins AL, Griffin CA, Li X, Jabs EW, and Uhl GR (1992) Human dopamine transporter gene (DAT1) maps to chromosome 5p15.3 and displays a VNTR. Genomics 14:1104-1106.

Varki A, and Altheide TK (2005) Comparing the human and chimpanzee genomes: Searching for needles in a haystack. Genome Research 15:1746–1758.

Vaughan CL (2003) Theories of bipedal walking: an odyssey. Journal of Biomechanics 36:513-523.

Verdin E, Dequiedt F, Kasler HG. (2003) Class II histone deacetylases: versatile regulators.Trends Genet. 19(5):286-93.

Vostrov AA, and Quitschke WW (1997) The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter: evidence for a role in transcriptional activation. . J Biol Chem. 272:33353-33359.

Walsh PS, Metzger DA, and Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. Biotechniques 10:506-513.

Wang C, Fu M, and Pestell RG (2004a) Histone acetylation/deacetylation as a regulator of cell cycle gene expression. Methods Mol Biol 241:207-216.

Wang E, Ding YC, Flodman P, Kidd JR, Kidd KK, Grady DL, Ryder OA, Spence MA, Swanson JM, and Moyzis RK (2004b) The genetic architecture of selection at the human dopamine receptor D4 (DRD4) gene locus. Am J Hum Genet 74:931-944.

Wang J, and Bannon MJ (2005) Sp1 and Sp3 activate transcription of the human dopamine transporter gene. Journal of Neurochemistry 93:474-482.

Wang QF, Prabhakar S, Chanan S, Cheng JF, Rubin EM, and Boffelli D (2007) Detection of weakly conserved ancestral mammalian regulatory sequences by primate comparisons. Genome Biol 8:R1.

Wasylyk B (1988) Enhancers and transcription factors in the control of gene expression. Biochim Biophys Acta 951:17-35.

Watanabe T, Naito E, Nakao N, Tei H, Yoshimura T, and Ebihara S (2007) Bimodal clock gene expression in mouse suprachiasmatic nucleus and peripheral tissues under a 7-hour light and 5-hour dark schedule. J Biol Rhythms 22:58-68.

Wei LN, Law PY, and Loh HH (2004) Post-transcriptional regulation of opioid receptors in the nervous system.Review. Front Biosci 1:1665-1679.

Wenderoth N, Debaere F, Sunaert S, and Swinnen SP (2005) The role of anterior cingulate cortex and precuneus in the coordination of motor behaviour. Eur J Neurosci 22:235-246.

Wendland JR, Hampe M, Newman TK, Syagailo Y, Meyer J, Schempp W, Timme A, Suomi SJ, and Lesch KP (2006a) Structural variation of the monoamine oxidase A gene promoter repeat polymorphism in nonhuman primates. Genes Brain Behav 5:40-45.

Wendland JR, Lesch KP, Newman TK, Timme A, Gachot-Neveu H, Thierry B, and Suomi SJ (2006b) Differential functional variability of serotonin transporter and monoamine oxidase a genes in macaque species displaying contrasting levels of aggression-related behavior. Behav Genet 36:163-172.

White RJ, and Jackson SP (1992) The TATA-binding protein: a central role in transcription by RNA polymerases I, II and III. Trends in Genetics 8:284-288.

Wray GA (2007) The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 8:206-216.

Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, and Romano LA (2003) The Evolution of Transcriptional Regulation in Eukaryotes.. Review. Mol. Biol. Evol. 20:1377-1419.

Wu D, Li T, Lu Z, Dai W, Xu M, and Lu L (2006) Effect of CTCF-Binding Motif on Regulation of PAX6 Transcription. Investigative Ophthalmology & Visual Science 47:2422-2429.

Zamorano R, Suchindran S, and Gainer JV (2006) 3'-Untranslated region of the type 2 bradykinin receptor is a potent regulator of gene expression. Am J Physiol Renal Physiol 290:456-464.

Zawel L, and Reinberg D (1993) Initiation of transcription by RNA polymerase II: a multi-step process. Prog Nucleic Acid Res Mol Biol 44:67-108.

Zhang J, Webb DM, and Podlaha O (2002) Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. Genetics 162:1825-1835.

Zhou Q, Kindlundh AM, Hallberg M, and Nyberg F (2004) The substance P (SP) heptapeptide fragment SP1-7 alters the density of dopamine receptors in rat brain mesocorticolimbic structures during morphine withdrawal. Peptides 25:1951-1957.

Zvonic S, Ptitsyn AA, Kilroy G, Wu X, Conrad SA, Scott LK, Guilak F, Pelled G, Gazit D, and Gimble JM (2007) Circadian oscillation of gene expression in murine calvarial bone. J Bone Miner Res 22:357-365.